# Modi Operandi of Social Network Dynamics

## The Effect of Context on Scientific Collaboration Networks

Julie Marcia Birkholz

# Modi Operandi of Social Network Dynamics

*Reading committee:*

prof. dr. Marshall Scott Poole
prof. dr. Maarten van Steen
prof. dr. Paul Wouters
prof. dr. Peter van der Sijde
dr. Chris Baerveldt


Cover image designed by Adam Jonas in Gephi
of a slice of VU Amsterdam Computer Science
researchers' collaboration patterns
Printed and bound CPI Koninlijke Wohrmann B.V.
ISBN: 978-94-6203-824-0
SIKS Dissertation Series No. 2015-12

vrije Universiteit amsterdam

SIKS

The work reported in this thesis has been carried out at the Vrije Universiteit Amsterdam under the auspices of the research school SIKS (School for Information and Knowledge Systems). The research was funded by the Network Institute, Vrije Universiteit Amsterdam and the Rathenau Institute, The Netherlands.

VRIJE UNIVERSITEIT

# Modi Operandi of Social Network Dynamics
## The Effect of Context on Scientific Collaboration Networks

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. F.A. van der Duyn Schouten,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Sociale Wetenschappen
op vrijdag 12 juni 2015 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Julie Marcia Birkholz

geboren te Menomonee Falls, Wisconsin, Verenigde Staten

promotoren: prof.dr. P. Groenewegen
prof.dr. J.M. Akkermans

# Acknowledgments

The writing of an academic piece is, relatively speaking, a science that requires practice and time. Eventually everything does fall into place and can be presented in some comprehensible format that can be judged whether deserving of a title of doctor; but writing the acknowledgments is an entire different project. The PhD project is an individual task, but it is not done entirely in isolation, it is here where I am able to acknowledge all those that contributed to this journey.

Thus, first and foremost, Radboud. I came to Amsterdam for you, you supported me through every single step of my Dutch journey, both the good and the bad, and I exited as a happy, married, mother, Dutch citizen, and now with a title of doctor! And you, of all people, continuously encouraged me to pursue a PhD; without you I would have never convinced myself to embark on this insane pursuit. In taking that step in 2009, a new chapter began that has garnered so many lessons and gifts. Among them of course is Caius! He entered our world at the end of the PhD project and perhaps threw all of the planning out the window. He also gave us both a necessary reset– where through new priority setting, I, in any case, brought clarity through baby-induced-exhaustion to organize the overall message of my research. I am so grateful for the continued support, patience, and space that you give me. One could not ask for a more patient and understanding partner, who without question does everything for our family. A very special thanks is also in order to you for suggesting the title of this dissertation. Radboud, thank you for your belief in me and my capabilities, as well as the project that is us! Without you none of this would be possible.

Of course, a giant thank you goes to my promoters- Peter Groenewegen and Hans Akkermans! Thank you for taking a risk and selecting me to pursue this research in the first place. I remember my interview like it was yesterday, regretting that I was too argumentative in suggesting that both networks on consequences (traditional SNA) and antecedents (use of SIENA and other models) needed to be used in investigating the research question I would be tasked with if selected. This suggestion lead to some discussion not only between myself and the inter-

viewers, but also between the three interviewers themselves. I remember coming home feeling deflated thinking I should have kept my big mouth shut and brainstorming that night with Radboud about different career options. To my delight, I received a call the day after from Peter that I had been selected! And shortly thereafter realized these discussions would become the cornerstone of our meetings, which transpired into countless research ideas. Peter and Hans, thank you for allowing me plenty of intellectual space to explore various research avenues and get lost in the interdisciplinary madness that was my PhD research.

Hans, I am extremely grateful for the candid conversations and support I received from you throughout my project. You are an exceptionally intelligent researcher, of which I am in awe. Your practical experience often led to so many wonderful anecdotes for research including the great story of developing ideas on random walks during the Mali road trips in the back of a bus. You showed me through example that research can literally be done everywhere, in any space of time– with a good idea, time and patience it will materialize! Hans, I am also grateful for your compassion during difficult times, especially after the birth of my son; I could not have asked for a better promoter who understood the importance of having things stable at home in order to concentrate on research. Thank you Hans for trusting in me and supporting my research and my research journey. I look forward to our future work on fields!

And to Peter (where do I start with Peter!) you were my daily supervisor, just a few doors down the hall from my office. You saw the best and the worst of my project. I am extremely grateful for the freedom you gave me to make my own mistakes and learn how to really conduct research. You whole-heartedly supported my collaborations with the many computer science researchers, for which I am grateful of your patience. Your involvement in many research projects is a testament to your great curiosity for knowledge, which you instill in your PhDs, thank you! I had the pleasure of visiting you during your sabbatical in the US and getting to know you as a friend while there. Although, it has taken me years to understand your ramblings, now that I don't hear them I miss them greatly. Peter, thank you; thank you for your honest, continued support and understanding, and giving me space during my leave to take care of my son. I would not be a post doc researcher today without your support and training that allowed me to develop myself as an independent researcher, thank you! I look forward to our many projects in the future.

This project started with a practical question: can an institute– a research center, an university, make a difference in the research that is being done by individual researchers. This dissertation is evidence of this exploration and the steps I undertook in attempting to comment on how context influences the success and emergence of social networks in investigating the publication patterns of Dutch Computer Science researchers. Although, this project could have not been realized without the input of Dutch Computer Science researchers. Thus, a special thank you to all those researchers that served as my research subjects. Those that I interviewed, which will remain nameless and anonymous as promised, as well as those that took the time to fill in my survey, thank you! I could not be here today without your insights, data, and feedback. I wish you all success with your continued work and I hope I have done some justice to describe the complexities of the field of Dutch Computer Science where you conduct your work.

Thank you to the members of the dissertation committee: Scott Poole, Maarten van Steen, Paul Wouters, Peter van der Sijde and Chris Baerveldt who invested time in providing feedback on this dissertation, which yielded a much better end product. A special thank you goes to Scott, your theory construction course aided in identifying the key theoretical contributions of my dissertation. I am honored that you were able to make the trip to Amsterdam to be present at my defense.

The Network Institute, Department of Organization Sciences, and the Rathenau Institute financially supported this project, for which I am grateful! Thank you, Edwin Horlings from the Rathenau Institute for your always insightful comments, enthusiasm, and interest in my research project. I am very grateful to the Network Institute which provided me with many opportunities to meet other researchers who were also interested in pursuing interdisciplinary research. Without this framework I would have never met Rena Bakhshi, Paul Groth, Shenghui Wang, Maarten van Steen, and Frank van Harmelen, among many others! My growing expertise is partly attributed to what I learned in pursuing projects with you all- a new language, a new way of thinking, and working! Thank you.

A very special acknowledgment needs to be extended to Rena Bakhshi– I could not have done this PhD project without you! In the fall of 2011 I was stuck and encountered Rena at a Network Institute event. I was searching for something that Rena seemed to have- a model to study network dynamics of large networks where I could consider external effects, the mean-field model! For a year we worked hard together experimenting with the model; finalizing our first paper together, over Skype from the US and Australia, on the morning of my wedding; we sent off the first of our many publications! Rena, there is not a better partner-in-research-crime than you! You always had the patience to explain to me the ins-and-outs of mathematical models and dared to question the silly traditions of social science. As my research found cohesive co-authorship groups reap success in environments like the VU, we are evidence of just that. Thank you for teaching me LaTeX! I am indebted to you, and so grateful to be both your co-author and friend and honored to have you as my honorary paranymph.

Many thanks are also due to the Department of Organization Sciences at the VU! First, a special thank you to Elles and Welmoed. You are both lovely and genuinely sweet people who are an awesome and immeasurable asset to the department! I do not have the space to thank all of my previous colleagues, but thank you all for the conversations at the coffee machine, for feedback in seminars, and for laughs over drinks, And of course the Metrogebouw N–329 peeps and neighbors, who came and went: Christine, Nicoletta, Daphne, Dirk, Judith, Pritha, Sandra, Priyanka, Bas, Cedric, and Isabelle. You made working life at the VU great! My SIKS buddies – Hendrik, Ivan, and Nana thanks for making nerdy computer science weekends in isolated Dutch villages fun! I am so grateful for the friendships that emerged from talks on networks, publishing, and general research madness! A special thank you to, Daphne, for being my friend and supporting me in navigating the roughest waters of my PhD. A special thank you to Dirk as I am convinced I would not be the researcher nor excellent karaoke singer that I am today without the time that I had to soak up your greatness!

And my dear friend and paranymph– Christine. I had the great pleasure of organizing two SNA workshops with you and we always joked we should have married each other during those workshops as we read each other's minds! I am beyond thankful for you as a co-author, roommate, co-mother, and friend! You held me up when shit hit the fan, called me even when I said I didn't want to be bothered, and held my hand both physically and mentally during both professional and personal hardships. I would not be who I am as a researcher or friend without you. Thank you for challenging me to be better. You are an amazing, intelligent researcher, which I look forward to sharing many more chapters of my life with. And also during the defense as it could not be more fitting that the one that has supported me daily throughout this research journey stands by side on this day! Thank you!

My second protector during the defense – Adam Jonas; I met you the very first day of my research project in October 2009, which just happened to be among your very first international social network conferences at the VU Amsterdam. I remember shaking your hand and striking up a conversation with you, which immediately lead to four days of non-stop outings in Amsterdam! Somehow we have been able to see each other every year since; staying on each others couches with our respective partners putting up with our network nerdy-ness! We, Radboud included, have sincerely enjoyed getting to know you over these years and although we may only physically be in each others' company once or twice a year we cherish the time we have with you! You are like family to us Adam and we look forward to our continued yearly shenanigans. Your radical nerdy-ness is contagious, as well as your laughter. You are an amazing, talented researcher and I look forward to keeping up on all the network nerdy-ness with you in the future. Thanks for the lovely rock-n'-roll Gephi network visualization that graces the cover of this book. Also, thank you for standing next to me during the defense, as you were part of this journey from beginning to end.

To my many other friends who have been witness to this PhD journey, thanks for putting up with it! Karin, thanks for being my friend and just listening and reflecting. Anita, thanks for helping me prepare the letter that got me this PhD project, and for being a friend, great neighbor, and an evening-time working buddy! Maryam, thank you for reminding me to be present and helping me find my 2012 anthem– I was here. To all of those that I have shared beers, dinners, laughs, and the much needed distraction from the PhD, thanks for reminding me that– *"Dit is de leven, En ik weet dat dit de leven is, En ik weet niet of het nog lang duurt, Maar zolang het er nog even is, Geniet ik er van, elke dag."* (Sef, 2011).

Of course none of this would have even been remotely possible without family. I am extremely grateful for the Reijn-van der Weij family, the Birkholz family, the Nordeen family and the Wachuta family. I love you all and words cannot describe my thankfulness for all of you; but I will try and say something to reflect how appreciative I am to you and your support for me in this PhD journey, in no particular order of appreciation!

The Reijn-van der Weij family, you all welcomed me to The Netherlands with open arms. I could not have asked for a more accepting family to marry into. Some notable moments though, in regards to the PhD: Joris, I could not have survived the end of the PhD without your visits to care for Caius that allowed me a few hours to write while you trained Caius to become addicted to ice cream! And Kunieke, for the always great advice on navigating my future career! Eep, your comic relief always puts a perspective on the things that really matter- not what we necessarily do from 9 -5! Steph, you are a shining example of balancing being a mama and working, it is an awesome task, which you do so well. Boudewijn and Willemijn you light up our lives! En Marjo– zonder jouw was mijn PhD onmogelijk geweest na de geboorte van Caius. Ik kon je altijd bellen, altijd vragen om te helpen, en je was er altijd. Voor ons, voor Caius, voor mij! Dank je wel dat je me altijd komt redden. Jij bent een van de sterkste vrouwen die ik ken. En vindt het een eer dat je mijn schoon-moeder bent!

The Birkholz family, thank you for supporting me during this project. My Dad (Barry) and JoAnn, thank you for always being interested in what I was doing and curious and rightfully so worried about where I was going! And of course for always knowing I had a home to come to, in case it didn't work out here across the ocean. And to my Dad, who taught us all to stand up for ourselves, thanks for letting me have a big mouth. Sarah and Eric, my little sis and bro, thanks for putting up with the absence of your sister at all important events in your lives, while I pursued an academic career abroad. I will try my hardest to make it to important events in the future. Even though I am never there, I will always be here if you need me! A very special thanks as well to my Grandma MaryAnn and Grandpa John B who always encouraged us to pursue education, and to speak our minds.

The Nordeen family– To Grandma Joanne, thank you for your continued support of everything I do and reminding me when things are hard that I should pull myself together and sit on the edge of my bed before I get up in the morning, like Grandpa Dan always did, and shout – "I feel terrific!". To my Uncle Mark Nordeen who assisted in helping me develop a number of procedures within MS Access to organize and analyze my data, and thus suffered through too many academic explanations of what I was doing. You are a great uncle and friend and I will always share a family fry with you. To Alex, thank you for reading my work and questioning me and continuing to question and challenge the systems that we have built as social beings, you are amazing! Keep fighting for yourself and your education! And Julie Nordeen, thanks for hosting us yearly, and of course loving and caring for Mark and Alex!

The Wachuta family– my Mom (Marcia) and Ritch, thank you! No words can express my gratefulness. Thank you for supporting me in every crazy journey that I told you I was going to pursue. Thanks Mom for always encouraging us to "have a great idea!". You both have had to suffer the heartbreak of the midnight crisis calls of a daughter who was gallivanting across the world in pursuit of her dreams and not being able to physically see me as often as you like. And as a mother, I know now that this is heartbreaking, you want nothing more than to hold your child and protect them from all pain. Without your encouragement already at a young age to leave a softball team when the coach told me I threw like a girl, or to come back from rejection from high school choir, and the comfort you gave me as a child, who always came home from sleepovers early homesick; I could not be the woman, nor mother I am today that has the certainty to pursue all my hopes and desires no matter where they take me. You laid the first stones that make me feel so wonderfully comfortable, certain, and positive about the trajectory of my life. Thank you!

And to my true side-kick since 1983– BJ, my brother, you are awesome! Even when you think what I am doing is nuts, you always support me (also financially:)! Thank you in elementary school for purposely failing the test to get into the gifted and talented program when I didn't get in, as you thought it was unfair; getting me through organic chemistry in high school; being at all the family things I wasn't and holding the fort down; fighting over principles; and so many, many late nights talking about everything and nothing. Only someone like you would fly across the world on a whim when I call you stressing about how on earth I will finish my PhD, pack up my apartment, prepare it for renters, move to Gent, get the house set up, take care of a then busy 18 month old, deal with my loss of leaving the great city of Amsterdam in a week's time! You are irreplaceable, you're my brother, my friend, and whatever journey I am on you are somehow there too. Thanks B, for everything!

To my curious, strong-headed, delightful son– Caius. May you always find the courage to pursue your dreams; even when it is a difficult task, some may doubt you and try to knock you down, but just keep going! And may you find happiness through all the dream-making, like I did with you.

Finally, again, Radboud, words are undeserving for my love and gratefulness for you, but given this is where these words are said- *dank je wel voor alles*, this book is for you and the journey that is us.

# Contents

# 1 Introduction

Social networks matter. The array of relationships within a social network allows individuals to tap into important resources (Burt, 2005; Coleman, 1988; Granovetter, 1985). It has been shown that the structure of the network, as well as the position of individuals within a social network, have consequences for outcomes (Knoke, 1990). However, optimal network structures remain debated (Burt, 2000; Coleman, 1988; Obstfeld, 2005).

Research suggests a number of individual- and contextual-level contingencies to this success (Battilana and Casciaro, 2012; Burt, 2000; Burt et al., 2000; Hansen et al., 2001; Podolny, 2001). Additionally, external factors are said to play a greater role in the success of networks of individuals than an individual's characteristics (Burt et al., 1998); thus questioning whether the conditions under which relationships are realized has an effect on its likely success. This is in contrast to the dominant assumption in network theory that individuals organize their networks and form ties to other people autonomously and relatively independent of any contextual influence (Burns and Stalker, 1961; Daft, 1982). Thus, I pose the following question in this dissertation research: *How do contextual factors influence social network structures?*

In this research, I identify how the consideration of exogenous context as a non-explanatory variable serves as a blind spot for understanding the emergence of network structures. I outline the existence of a theoretical gap in network theory given the dominance of the individual as a driver of networks in network theory. This has ontological implications that further retain context as a blind spot in the empirical investigation of the effect of context on network structures. To pursue this research, I present a theoretical framework that integrates network and structuration theories to investigate the interplay of context – as measured by the rules in the environment – and individual and endogenous factors as resources to explain the success and emergence of network structures. This framework considers contextual factors as determinants which explain an individual's network structures. A mixed-methods network approach is taken for the evaluation of the effect of contextual factors in explaining both antecedents and consequences of network structures. This question is explored in an academic setting, where the scientific collaboration networks of Dutch Computer Science researchers are investigated. I address this question in separate empirical studies presented in three chapters. In this introductory chapter current theory is reviewed and a theoretical framework is presented (given gaps in theory), following the research design and the methodology used to investigate context in explaining the *modi operandi* of networks.

## 1.1 Social Networks

A social network represents relations between actors (Borgatti et al., 2009). The study of social networks within social sciences, as this dissertation reflects, has its roots in sociological theories where relationships form a part of the basis for understanding behavior (Durkheim, 1951; Simmel, 1955); where all actions are embedded in networks (Granovetter, 1985). (It is not the goal of this research to reflect on these historical roots, for a review see (Wasserman and Faust, 1994).) These relations – a set of ties between nodes (entities) – define a network. These networks reflect a type of relation (e.g., a friendship tie in a friendship network or advice tie in an advice network).

Networks can be represented as graphs where positions and structures are systematically analyzed (Wasserman and Faust, 1994). These principles originate from graph theory (van Steen, 2010), which provides mathematical descriptions of characteristics. These measures serve as a tool for social scientists who research social networks to systematically investigate antecedents and consequences of network structures that provide insight into mechanisms and social processes. Networks, therefore, can be studied by examining structure, content or function.

The structures of these relations, of interest to this research, are seen as proxies for understanding social structure (Blau, 1974; Burt, 1980; Coleman, 1988). The study of the structure of these relations compromise three domains: the study of the network itself, antecedents and consequences of networks.

### 1.1.1 Consequences of Social Networks

The study of network consequences comprises the largest area of current work in the social sciences; the aim of such studies is to explain how positions or structures lead to variations in outcomes (Borgatti and Foster, 2003). Related to this is the way network structure, as well as the positions of individuals within social networks, has consequences for outcomes (Ahuja, 2000; Burt, 2005; Coleman, 1988; Cummings and Cross, 2003; Fleming et al., 2007; Knoke, 1990; Obstfeld, 2005; Tortoriello and Krackhardt, 2010). In this research I concentrate on network structures.

Scholars have long debated the benefits of particular network structures (Burt, 2005; Coleman, 1988; Granovetter, 1985), suggesting differences in the access granted to valued resources (Brass, 1984; Ibarra and Andrews, 1993). This access is related to constraint: the extent to which an actor's alters are connected to each other. Constraint is considered from two extremes : structural holes (i.e., no or low constraint) and closure (i.e., high constraint). A structural hole exists when an actor is situated between other actors who are not linked. Such network structures expose the ego to non-redundant information, which facilitates brokerage (Burt, 2004). Inversely, high constraint or closure describes a dense, cohesive network structure (Coleman, 1988) that mobilizes resources and people through multiple connections, where all actors have the ability to share and exchange knowledge (Coleman, 1988; Granovetter, 1985; Tortoriello and Krackhardt, 2010). Both network structures are seen to have advantages. Brokerage (e.g. low constraint) provides advantages for the individual due to increased control and access to information (Fleming et al., 2007; Mehra et al., 2001; Obstfeld et al., 2014). Dense networks (high constraint) facilitate performance through increased identification with a group (Borgatti et al., 2009) and trust (Borgatti and Foster, 2003).

Recent work has suggested that individual and contextual conditions play a role in the performance of networks. Networks that are low in constraint facilitate the performance of non-complex, less knowledge-intensive tasks (Hansen, 1999; Krackhardt, 1992). Inversely, high constraint in networks has been shown to be most effective for completing complex knowledge tasks (Cummings and Cross, 2003; Reagans and McEvily, 2003). Burt (2000) suggests that under conditions of high uncertainty and in small groups, low constraint (i.e., structural holes) is most beneficial for performance, while a constrained network structure has been found to improve performance when work is done under low uncertainty in larger groups. Similar findings were confirmed by Battilana and Casciaro (2012), suggesting that change agent networks are more successful in cohesive networks with higher constraint in organizations that undergo changes that divert largely from the status-quo. Networks with less redundant contact perform better in new markets or with new technologies of uncertainty than in familiar terrain where redundant contacts prove more instrumental to eventual success (Burt, 2000; Hansen et al., 2001; Podolny, 2001). Burt et al. (2000) found differences in the success of the networks of managers in American and French companies, suggesting that culture as an external factor plays a role. External factors are said to play a greater role in networking decisions of individuals than the role of an individual's characteristics (Burt et al., 1998).

Thus, empirical work on network consequences suggests that specific contextual conditions under which networks are invoked influence the success of different structures. However, knowledge of the outcomes of these networks is incomplete without a comprehension of how network structures emerge (Salancik, 1995).

### 1.1.2 Antecedents of Social Networks

Understanding network antecedents provides insight into the conditions under which specific structures emerge by way of relationships between actors. A number of generative mechanisms have been identified as antecedents that lead to network structures (Monge and Contractor, 2003). A social network is said to emerge through the creation, maintenance and dissolution of relations between actors. The roots of these structures describe the origin of these mechanisms.[a] The roles these generative roots play in the generation of network structures remain debated, with theories suggesting different *modi operandi* (Ahuja et al., 2012; Liben-Nowell and Kleinberg, 2007; Shumate and Contractor, 2013; Whitbred et al., 2011).

Theories on the emergence of social network structures suggest that multiple generative mechanisms govern network formation (Monge and Contractor, 2003). (See Table 1.1, Examples of Generative Network Mechanisms, for a review; this is not an exhaustive list, but rather a descriptive guide of mechanisms related to specific roots. Some key mechanisms are discussed below and in specific chapters.) Following Lusher et al. (2012), I conceptualize these generative mechanisms as having specific roots, which can be categorized by their origin and are identified as:

(1) endogenous network factors;

(2) individual factors; and

(3) exogenous contextual factors.

**Endogenous Network Factors** Endogenous network studies have resulted in one of the largest bodies of empirical work on network dynamics in physics, computer science and the social sciences.[b] These mechanisms have been detected in numerous types of networks; this internal basis for associations demonstrates how their roots are endogenous. The prediction of the network structure, and the positions of individual actors based solely on previous structures and positions, has largely been sourced from mathematics and physics. These are endogenous network mechanisms[c] and are inherent in the organization of the network.

---

[a] I refrain from referring to these roots of mechanisms strictly as origins, as employing such a term assumes a type of evolutionary theory of these processes.

[b] In the hard sciences this work is referred to as the *link-prediction problem* (for a review, see Liben-Nowell and Kleinberg (2007)).

[c] In the field of social network studies, endogenous network factors are often referred to as ”self-organization” (Lusher et al., 2012), but self-organization serves as an umbrella term for network dynamics, specifically in the natural sciences. Given the interdisciplinary nature of this dissertation research, I have elected to not use the term self-organization but ”endogenous network factors” instead.

Table 1.1: Examples of Generative Network Mechanisms Classified by Roots

| Endogenous network factors | Individual factors | Exogenous contextual factors |
|---|---|---|
| ***Preferential attachment*** relationships emerge among those actors who have many relationships and therefore attract an increasing number of relations (a power law) (Adamic and Huberman, 2000; Albert and Barabási, 2002; Price, 1976).<br><br>***Transitivity*** relationships emerge to establish a cluster among three individuals (Simmel and Wolff, 1950). (See also cliques and cohesive groups (Newman and Park, 2003; Wasserman and Faust, 1994)).<br><br>***Closure*** relationships emerge based on structural balance (Cartwright and Harary, 1956), with tendencies to invoke a relationship pattern to remain a focal actor (brokerage) (Burt, 1992) versus establishing closed, cohesive relationships (Granovetter, 1973).<br><br>***Reciprocity*** relationships emerge based on the tendency to exchange and/or reciprocate relations (Blau, 1964). | ***Homophily*** relationships emerge based on similarity of traits (Blau, 1977; Ibarra, 1992; McPherson et al., 2001).<br><br>***Exchange*** relationships emerge based on reciprocal exchange of information or material resources, sometimes dependency exchange (Klein et al., 2004).<br><br>***Self-interest*** relationships emerge in favor of personal preferences (Mehra et al., 2006, 2001; Shah, 1998).<br><br>***Collective action*** relationships emerge based on the pursuit of a common goal (Taylor and Doerfel, 2003).<br><br>***Cognitive mechanism*** relationships emerge influenced by the perception of attitudes toward each other (Borgatti and Foster, 2003; Krackhardt, 1987; Palazzolo et al., 2006).<br><br>***Balance theories*** relationships emerge based on maintaining similar relations to others in the same position (Khanafiah and Situngkir, 2004). | ***Proximity*** relationships emerge based on co-location (Van den Bulte and Moenaert, 1998). |

A number of mechanisms are attributed to the endogenous (or internal) root, three of which are highlighted here. Scale-free networks, exchange and transitivity are the most relevant to this research, as they provide insight into reciprocity and relationship ties, two key issues in social networking. Scale-free networks (Barabási and Albert, 1999; Price, 1965) can be explained through the mechanism of preferential attachment or so-called power laws. Here, the growth of the network occurs through a process in which a certain quantity of ties or relations are distributed among individuals, according to how much they already have, resulting in cumulative advantage – where 'the rich get richer'. Exchange is a network mechanism that explains the tendency of exchange as represented through a reciprocal relation between two actors (Blau, 1964). An additional mechanism is transitivity, this explains the tendencies among three or more actors to connect based on ties between one or more actors' previous relations (Simmel and Wolff, 1950). Recent reviews of the effects of endogenous networks on social networking has questioned the strength of these mechanisms in the *modus operandi* of networks as the sole predictor of network emergence (Liben-Nowell and Kleinberg, 2007). As a result, the influence of non-endogenous mechanisms has garnered increased attention.

**Individual Factors** A second stream of research, largely within the social sciences, examines the effect of individual characteristics on network dynamics. This perspective is distinct from work in endogenous network mechanisms as it recognizes a possible dual role of endogenous network mechanisms and the characteristics of the individual that lead to this network emergence. These characteristics include traits, attitudes, dyadic relationships and formal positions. These mechanisms explain network dynamics through the consideration of the characteristics, attributes or attitudes of the individual members of the network.

In the social sciences, the individual plays a leading role in conditioning the relationships within the social network. This perspective can be partly explained by prominent views in the social sciences over the past 30 years, where there has been a tendency to understand higher level social phenomena based on explanations anchored in the individual and his or her traits and/or characteristics. Thus, networks are understood as relational processes where individuals create, maintain or dissolve a relationship within an embedded system.

A number of mechanisms have been attributed to individual characteristics. For example, homophily describes the process through which ties are established based on the common attributes of the actors (Blau, 1977; Ibarra, 1992; McPherson et al., 2001). The qualities of individuals also play a role in the likelihood for ties to emerge (Kilduff and Krackhardt, 2008). Self-interest invokes the establishing, maintaining or terminating of relationships due to a personal preference or desire (Mehra et al., 2001; Shah, 1998). Balance theory suggests a mechanism in which actors aim to maintain consistency in relations with similar actors based on characteristics (Monge and Contractor, 2003). Personal qualities and unique traits and behaviors are central to these individual mechanisms.

**Exogenous Contextual Factors** In addition to the internal and individual aspects of these two types of generative network mechanisms, context is also a factor in considering dynamics. Exogenous contextual factors are mechanisms such as physical proximity, which leads to interaction and thus potential relationships (Borgatti and Cross, 2003). This proximity has been found in the study of networks comprised of members from the same national or regional culture (Monge and Eisenberg, 1987; Owen-Smith et al., 2002), as well as in studies of organizational structure, working conditions and demands (Balkundi et al., 2007; Burkhardt and Brass, 1990; Danowski and Edison-Swift, 1985; Shah, 2000; Tichy and Fombrun, 1979). Exogenous contextual factors are most often thought of as dyadic (a relationship based on a common affiliation between two actors). Environment and circumstances, whether between similar or different partners, are significant elements for the establishment of networks. However, given the epistemological starting point of social network theory with the individual as driver, exogenous contextual factors are not conceptually considered as explanatory variables on the network trajectories of individuals.

## 1.2 Social Network Dynamics[d]

While a significant amount of empirical and theoretical work has contributed to our understanding of network mechanisms, less understood is the way that these generative roots — the *modi operandi* of network dynamics — interact in explaining emerging network structures. By specifying how these interactions explain different network structures, one can define the conditions under which specific network structures are likely to emerge, in order to describe the dynamics of social networks. At least four recent theories exist that attempt to explain the *modi operandi* of social networks.[e]

These theories take two distinct directions, based upon the role of the generative roots. I briefly review these theories here and organize them in respect to the role of the roots of the mechanisms, as summarized in Table 1.2, Role

[d]In the natural sciences, this may also be considered network evolution, self-organization, or the link-prediction problem. I maintain the use of Social Science terms in this dissertation research.

[e]Emerging work on network dynamics suggests that content (i.e., semantics) should also be considered in addressing the emergence of network structures and positions (see Carley (1997)). The majority of this work deals with the co-evolution of social and semantic networks, e.g. mind maps (Carley and Palmquist, 1992; Danowski, 1993; Diesner and Carley, 2005; Lehmann, 1992; Popping, 2003), although it remains in its infancy. The effect of semantic networks on social networks was also explored in this research, although it is not presented in this dissertation as it did not serve to identify the determinacy of exogenous contextual factors. A number of exploratory studies were accomplished that make qualitative contributions to how semantics and network positions interact to explain network structures, cf. (Moser et al., 2013b) and a related presentation (Moser et al., 2013a).

of Generative Roots in Theories of Social Network Dynamics. I classify the four theories as having two distinct roles: a leading role and a dependency role.

### 1.2.1 Theories Proposing the Leading Roles of Roots

In the first of these four theories, generative network mechanisms are thought to be contingent on the type of relationship the members have (Shumate and Contractor, 2013). These network consequence studies also examine how the type of relation has diverse implications for outcomes. Seven types of networks have been identified: flow, affinity, representational, semantic, technological, physical, and affiliation (Shumate and Contractor, 2013). Each type invokes a specific root (or set of roots) that relates to the likelihood of the emergence of different sets of mechanisms. For example, collaboration is an affinity-type network. It implies flow of information or resources and an exchange of information and/or knowledge, as well as a strategy by actors. In these networks, individual characteristics are said to drive network dynamics. A second theory proposes that mechanisms are dependent upon the development stage of the network (Poole and Contractor, 2011). This theory was specifically developed to understand team networking processes, suggesting that all roots play a possible role and the dominant roots are dependent on the development phase.

### 1.2.2 Theories Proposing the Dependencies of Roots

The third and fourth theories identify a starting point for network dynamics. Ahuja et al. (2012) purports a co-evolutionary process that is driven by the decisions of individual actors to form a relation with others. This theory is distinct from the theory of Poole and Contractor (2011), which purports that micro-dynamics (e.g., sharing an attribute) dominate the evolution process. Another theory of this type suggests individual characteristics (conceptualized in an encompassing term as mechanisms of internal factors) play a key role in explaining network structures (Whitbred et al., 2011). This theory coincides with the large empirical work on network dynamics in the natural sciences that has purported that endogenous network factors largely explain dynamics, with recent work suggesting a stronger role for the inclusion of individual characteristics in predicting relation emergence (Liben-Nowell and Kleinberg, 2007). These four theories and the two directions they represent are shown in Table 1.2, where the three previously outlined generative roots are also identified.

As outlined here, the antecedents of network structures can be explained through the different roles of generative roots. One set of theories suggest mechanisms with a specific generative root dominate the *modus operandi* of social network emergence (see Leading Role in Table 1.2). A second set of theories suggests conditions under which the effect of the roots is dependent on either the network type or the stage of development of the network (See Dependency

Table 1.2: Role of Generative Roots in Social Network Dynamics Theories

| Roots \ Theories | Leading Role (Ahuja et al., 2012) & (Whitbred et al., 2011) | Dependency Types (Shumate and Contractor, 2013) & Developmental Stage (Poole and Contractor, 2011) |
|---|---|---|
| Endogenous network factors | ✓ | ✓ |
| Actor characteristics | ✓ | ✓ |
| Exogenous contextual factors | | ✓ |

Types in Table 1.2). In all four theories the role of the individual as a driver of these dynamics is implicit.

Due to the dominant assumption that individuals organize networks and form ties to other people autonomously and relatively independent of any contextual influence (Burns and Stalker, 1961; Daft, 1982), it is no surprise that current theories do not explicitly clarify how exogenous contextual factors serve as possible determinants in explaining the emergence of structures. These theoretical assumptions create a limitation for investigating the potential effect of context in explaining network structures. For example, the empirical investigation of these networks is largely studied as segments of a population, i.e., a specific industry (Uzzi, 1997), a classroom (Knecht et al., 2010), or a disciplinary field (Owen-Smith et al., 2002). Keeping the study of network mechanisms bound to a single context absolves network researchers from investigating possible effects from the context.

### 1.2.3 Considering context

Thus, despite increasing empirical evidence that context serves as a detriment to the success of structures, context as an explanatory factor in the role of context as an antecedent to network structures is understated in current theory and thus overlooked in empirical studies. This is a blind spot in current network studies. Given this current paradigm, there is a need for further empirical evidence that exogenous contextual factors play a role in the network structures that emerge and the social selection mechanisms that explain them. To investigate this, within the limits of organizational theory in which this research is embedded, I sought a framework that explicates context to explain these networks, where networks could be seen as organizational behavior in the sense of a measure or outcome of organizing as a practice.

A number of perspectives exist within organizational theory to explain the effect of the organizational level on organizational structures, emphasizing the role of the formal organization. These include: institutional theory (North, 1991; Scott, 1995), resource dependence (Pfeffer and Salancik, 2003), and transaction cost dynamics (Teece et al., 1997). It is not my goal here to present these theories and their conceptual distinctions in depth, but rather to acknowledge that they play a role in how exogenous contextual factors explain structures. As an example of this perspective, I outline institutional theory. Institutional theory purports that institutions are socially constructed and provide guidelines for action through the repeated actions of members (Meyer and Rowan, 1991). Indeed, institutions are comprised of a "web of values, norms, rules, beliefs, and taken-for-granted assumptions" (Barley and Tolbert, 1997, pg. 93) whose rules serve as the blueprint for procedures and practices (Meyer and Rowan, 1977). These theories provide the logic necessary to understand how formal organizations govern the emergence of organizational structures. However, these consider the institution as a sum of its parts, where the interplay of context in relation to the individual, and that person's relation to others in describing observed behaviors, is overlooked (Barley and Tolbert, 1997).

In an attempt to rectify the relationship between the context and the individual, a number of extended studies have emerged from these theories. DiMaggio (1988) describes a limit to the constraint of the formal organization, suggesting that institutions do not completely determine the actions of individuals, as largely assumed in conceptualization in previous theories. More recent theories on institutional work (Lawrence et al., 2011) and institutional entrepreneurship (Garud et al., 2007) have sought to further illustrate the role of the individual in understanding organizational behaviors. Even within these theories, however, actions of individual are still portrayed as embedded within the institution.

Institutional theory as an example of this perspective sheds light on how organizations serve to constrain or facilitate different organizational behaviors, although it does not specify the roles of the three generative networks roots described in Section 1.2 above. Thus, organization theory, in juxtaposition to network theory, largely sees organizational processes as institutionally driven. As a result conceptually and empirically disparate bodies of literature on organizations (context) and individual networking behaviors (Ahuja et al., 2012; McEvily et al., 2014) have developed. Given that the possible interplay of exogenous contextual factors and other factors as a way of explaining emergent network structures is overlooked, I seek an overarching perspective that provides an explanation for the interplay of context and the individual in particular.

## 1.3 Theoretical Framework

Following the work of Desanctis and Poole (1994) and Whitbred et al. (2011), I draw from structuration (Giddens, 1984) to specify this interplay. In network studies, structuration has largely been employed as a framework for explaining the natures of structure and process in network emergence (Whitbred et al., 2011). I apply this framework to specifically explicate the role of generative roots in explaining the differences in network structure emergence. In implementing this theory, I do not have the same aim as Giddens in explaining the effects of institutional processes as they relate to motivations; rather, such a theory provides an explanation social network structure emergence. I explain here first generally the theory of structuration and then outline how I attribute the key concepts to current knowledge on the three generative roots of network structures.

### 1.3.1 Structuration

Structuration describes a dynamic social process that explains order within the system of society (Giddens, 1984). This system is under a constant state of rectification, where actors internalize the constraints of rules, given their resources, and take action through practices that lead to an established system. This is then replicated or revised by others and becomes normative for further action within an institution (Barley and Tolbert, 1997). Within this framework, the individual is a knowledgeable actor who is aware of the conditions under which actions are taken and creates structures in exercise of this agency, which then constrains them. Actors have agency in the sense that they reinforce and potentially alter the ways in which the system emerges from specific rules and resources.

Unlike network theories which present network structure as the outcome in using structuration (Desanctis and Poole, 1994; Whitbred et al., 2011), Giddens assumes that the (network) structure is part of this interplay, as the previous structure of relationships is a dependency on the outcome of structuration. Few works to my knowledge have further specified this distinction, but rather positioned the network structure as an outcome. This corresponds with current thinking in network theory that sees networks as the outcome of processes. Given that I specifically investigate dynamics, this distinction is important as it infers that previous structures influence future behaviors.

In particular, the concept of *duality* in structuration provides specification about the relationship between exogenous contextual factors and other network roots. *Duality* is how the actions of today influence tomorrow's through both exogenous and endogenous factors (Barley and Tolbert, 1997). It is comprised of rules and resources. Rules are one of the principles that structure practice, often initiated by an external body, but reproduced and reinforced through actors' actions, given their set of resources.

Rules serve as the blueprint for procedures and practices (Meyer and Rowan, 1977). These rules entail explicit goals, motivation statements that potentially influence the practice that actors employ in their work within an institution (Sewell Jr, 1992). Every organization has rules, whether implicit or explicit. These rules serve as a set of conditions that outline the limitations within a given trajectory. For example, a rule may be a policy guiding a process for undertaking a specific task that would potentially alter the way in which individuals approach the task (e.g. a policy on pay scales, or travel) and, thus, the potential relationship sustained in this process. Recent work has suggested that conditions of uncertainty and certainty are an explanatory factor in understanding the success of different network structures (Battilana and Casciaro, 2012; Burt, 2004). Policies that aim to provide certainty facilitate the building of particular social relations realized in a (network) structure. Case in point: a policy outlining the steps towards gaining a promotion aids in specifying the precise evaluation mechanisms that propel individuals to a specific end. A rule in the form of a policy creates a heightened awareness of a specific outcome, which further stimulates a specific set of behaviors on behalf of an individual to navigate within these boundaries.

If rules enable different behaviors, resources depict the means by which an individual enacts certain behaviors; they are "anything that can serve as a source of power in social interactions" (Sewell Jr, 1992). These can be tangible, such as money, and intangible, such as status. Resources are implicitly unevenly distributed, where one person has more resources than another due to knowledge or experience. Individual characteristics are most often recognizable attributes such as gender, formal status, interests, and affiliations. In addition, these resources can also be less overt, such as power or informal positions. Individual characteristics explain tendencies among individuals to employ certain behaviors, given the set of rules.

### 1.3.2 Structuration and Generative Roots

The concept of *duality* allows us to identify antecedents to the realized structures that individuals employ, given the constraining or facilitating role of rules and resources. This conceptualization mirrors present knowledge of network structures, where multiple roots stipulate possible combinations of outcomes measured through the structure of interactions. I delineate here how the three generative network roots can be conceptualized within the concept of *duality*.

Rules in this framework are conditions defined by an organizational space – such as an organization, a field, or a classroom – that binds or conditions actions through policies. These rules stipulate the possibilities of certain behaviors. Within networks, these can be considered generative roots that are exogenous. Through the identification of different policies or practices of a formal organizational body, one is able to distinguish differences in contextual settings.

Resources are the means which an individual uses to navigate the rules. Consequently, the attributes of individuals based on the dimensions of their personal characteristics, as well as the characteristics of the social relations in which they are embedded, are considered to be resources that influence the potential of organizational patterns. In regards to generative roots, these are endogenous network factors and individual characteristics. Endogenous network factors, as measured by network characteristics such as positions or structures, garner capital for the individual by way of resources (Burt, 2000; Coleman, 1988; Granovetter, 1985). They function similar to an intangible resource. In this conceptualization, previous network characteristics, as inferred in structuration, play a role in explaining the present and/or future actions of individuals. In addition, individual characteristics provide a measure for a number of attributes that also identify resources (i.e. power from a formal position, wealth from ownership or affiliation, age as a form of status or expertise). These two generative roots provide dimensions for resources.

Thus, each of these generative roots are considered in *duality* where an interplay occurs between rules and resources to realize structures (see Table 1.3). The combination of these rules and resources provides possibilities for a specific network structure to realize an action; technically speaking, therefore, rules are a determinant. This provides a framework to explore a variety of mechanisms where specific combinations of rules and resources lead to different structures and even similar outcomes.

Table 1.3: Conceptualization of Theoretical Framework

| Term in structuration | Generative Roots |
|---|---|
| rules | exogenous contextual factors |
| resources | endogenous contextual factors<br>individual characteristics |

For example, an organization with a specific set of rules (exogenous contextual factors) stipulates the conditions necessary to achieve promotion. In order to understand the ways in which individuals take action to pursue this promotion (assuming all actors are rational), we consider both the given resources of the individual (individual characteristics and endogenous factors) and the rules. These rules serve as a factor that specifies the conditions for likely emergence of a specific behavior. In networks, this implies that the existence of specific factors within the exogenous context enhance the likelihood of the emergence of specific network structures through the manner in which they facilitate different generative roots. To illustrate this point, we can consider a company that has a policy that provides incentives to individuals who work with others outside of their expertise. This leads individuals to foster relations based on individual characteristics that lead to heterophilous social relations. This can be compared to a company whose polices of homophily – tendencies to relate with those similar to you – may be attributed to the network processes. Thus rules serve to condition the likelihood of resource allocation and thus alter the mechanisms invoked to achieve a network structure.

Consequently, this framework implies that exogenous contextual factors serve as a determinant to network emergence. Previous network structures are incorporated in this framework as an aspect of the process of *duality* that yields the outcome (following previous network studies) of future network structures. The application of structuration as a framework has a number of advantages. First, it provides an explication for the three generative roots, suggesting an interplay where rules act as a determinant in explaining network emergence. Secondly, it explicitly infers dynamics. In an effort to expand on current application in network theory, I suggest that previous network characteristics are conceptualized as a dimension of resources in explaining future structures. Third, it provides a parsimonious explanation for the organization of the rich and growing literature on social network dynamics in the development of a clear set of propositions for considering exogenous contextual factors.

In identifying the relationships of these roots as an explanation of network structure, I imply a specific causal relationship which allows me to appropriate conditions for investigating the *modus operandi* of social network dynamics, where context serves as a determinant. In order to investigate this proposition, I research the ways that rules (exogenous contextual factors) influence the emergence of social network structures. Through this framework I am able to advance knowledge in the *modi operandi* by specifying the role of contextual factors to explain the diverse network structures observed. The following section on Methodology describes the steps I took to investigate this interplay.

## 1.4 Methodology

In this research, I investigate the effect of exogenous contextual factors as a determinant to the emergence of network structures. I review here first the various methodological approaches that are appropriate for investigating networks and, specifically, antecedents. I then outline the research design employed for testing a network model as a way to explain the role of exogenous contextual factors as a moderating influence on network emergence and success. Finally, I briefly outline the setting of this study of academic science, specifically investigating the scientific collaboration networks of Dutch Computer Science researchers. This setting is presented more in detail in Chapter 2,

### 1.4.1 Social Network Analysis Methods

In this research, I take the perspective that a network represents one type of relationship between individuals. As represented by a node in this case, it specifically comprises the individual researcher and a relation – a link or tie that represents a scientific collaboration relationship undertaken to publish an academic publication. These networks can be considered both ego networks and closed networks. Ego networks indicate the relationships that an individual has with others (alters) in their network. Closed networks assume that these relations are bounded by a context e.g. networks represent the relations of a school, or an organization. This perspective allows for a number of possible methods to test the exogenous contextual factors as an interaction term. In considering an appropriate research design, I first outline the social network analysis techniques used within the social sciences to investigate social network dynamics and then present how I will consider the investigation of the effects of exogenous contextual factors.

Network methods are relevant to the adopted design because they are distinct from traditional statistical methods since they concern interdependencies between individuals. These networks can be considered to be either ego networks or closed networks. Social network analysis techniques are used to identify positions of individual nodes and overall network structures. Often, results from social network analysis are combined with differential statistical analysis to test effects.

For studying networks, the emergence of network data for multiple time periods or network panel data needs to be considered. A large body of work in the social sciences employs the qualitative comparison of network structures for two or more sets of actors or networks (see Faust and Skvoretz (2002) for a review). Advances in knowledge regarding social network dynamics have also brought an increasing complexity of techniques for their investigation. I focus here, however, solely on the most commonly used computational methods for investigating the network dynamics implemented in the social sciences. Many network models implicitly consider the role of an exogenous contextual factor. Specifically, the algorithms used to model dynamics originating in the natural sciences conceptualize the dynamics of entities (e.g., atoms, molecules, cells and the like) that have the capacity to interact with other entities, where behaviors are constrained by conditions within the environment or physical space in which they reside. These conditions steer the trajectories of individual entities, thus giving rise to a (fixed) set of interaction options among the known possible interactions between two entities. Consequently, dynamic network models implicitly assume that context is a contingency for the emergence of different network structures. While the large majority of these network models are also used in the Social Sciences, the theoretical models developed to investigate social network structures ignore context as an explanatory variable because they consider small or boundary-driven networks.

These social network methods explain the likelihood of relations between actors through statistical models largely by use of exponential logarithms. However, it should be noted that these are exponential distributions based on the log of the ratio of probabilities for network relations. Given the knowledge domain for this dissertation, I do not aim to specify the mathematics or algorithms of these methods, but rather the theoretical implications that such assumptions bring in investigating dynamics.

### 1.4.2 Techniques for the Study of Social Network Dynamics

I review here the three main techniques commonly used in studies of social network dynamics, that is:

(1) Quadratic Assignment Procedure (QAPs) (Martin, 1999);

(2) multiple network modeling using ERGM (p* models) (Robins et al., 2007); and

(3) the SIENA model (Snijders et al., 2010).

These methods are often seen as in competition with one another for methodological dominance, yet in reality each of the models has distinct theoretical assumptions that influence the predicted emergence of the network. These distinctions occur at the level of unit of analysis, as either tie- or actor-oriented models. Tie-oriented models seek to predict the likelihood of tie emergence based on the fact that new ties emerge as a result of the existence of previous ties, while actor-oriented models give priority to the individual in investigating network emergence (Lusher et al., 2012). In addition, models were also identified according to whether they might consider moderation of the exogenous contextual factor as a means for explaining an outcome. The three main techniques are described more in detail below.

Quadratic Assignment Procedure (QAP) is devised for tie prediction based on the influence of the attributes of individuals or pairs of individuals on structural outcomes (Martin, 1999). Structural elements are measured over time and evaluated through a regression (considering network interdependencies). This method is most often used in questions of multiplexity, which seek to explain dynamics through an understanding of reinforced relationships. Consequently, for QAP, any type of multi-rooted or multi-level data, as considered in this study, is a limitation because it means mixing up different distributions.

A second model is ERGM (i.e., the p* model), which simulates the likelihood of tie formation of multiple network structures (Robins et al., 2007). The p* model uses evolving random graph models to identify structural patterns that signify that a specific mechanism is taking place. This method is used largely to consider single cross-sectional data sets. The p* model does not consider the order of networks of relations; rather, the same value is given to all networks in identifying network mechanisms. In these models, exogenous contextual factors are most often considered as dyadic parameters/variables (e.g., two nodes sharing an affiliation) (Lusher et al., 2012).

An extension of the p* model allows for the comparison of structural tendencies, which could identify different statistically significant patterns of generative mechanisms. In these models, coefficients are normalized to delineate the differences significant mechanisms (Robins et al., 2007). The benefits of this model are that mechanisms can be identified from the cross-sectional data of just one capture of a network. Given a good fit for the compared models, this model provides valid certainty through statistics of the differences. Considering the theoretical positioning of this research, in which the individual is constrained by an exogenous factor that either inhibits or facilitates particular network behaviors, as well as the presence of certain network factors, a tie-based model and, specifically, a model that does not explicitly consider previous states in modeling likelihoods of tie emergence such as ERGMs, is not particular suited.

The third most commonly used method in social network analysis dynamics studies is the use of the Simulation Investigation for Empirical Network Analysis (SIENA). SIENA provides a model for the investigation of dynamics in actor-oriented networks (Snijders et al., 2010) and is suited to closed networks that simulate the likelihood of network change based on: data on relations of individuals within the closed network from two or more time periods; and a set of variables, parameters and/or covariates that theoretically influences a possible change in the network. This method simulates all possible networks from the first given period – assuming that every individual has the potential to make a tie with the other individuals in the network – and then calculates the likelihood that the given parameters influence the change observed in the second (and/or consecutive) networks using Markov chain modeling (Ripley et al., 2011). These models are also inherently dynamic, as they consider previous states (or network structures), as well as a set of additional factors modeled in simulating the likelihood of future network states.

SIENA models are suited for relatively small (approximately $1,000$ nodes) closed networks (where all individuals know of one another and of each others' network behaviors). They are a limitation for considering and/or comparing bounded networks and investigating how exogenous contextual factors influence the emergence of generative mechanisms to explain structures, given the large state space needed to calculate the Markov chain model. Additionally, the statistical comparison of models through normalization of coefficients, as done in the p* models, is not applicable to SIENA due to the fact that in SIENA, statistics are inferred from the distinct starting points of these models (Lusher et al., 2012). A number of mathematical solutions exist for comparing Markov chains (Müller and Stoyan, 2002), although they have not been applied in social science studies due to the sensitivities of these simulations for starting data. A methodological extension of SIENA would be necessary to model network dynamics and consider exogenous contextual factors as a moderator.

Consequently, there are a number of techniques available for implementing social network analysis pending the research in question. Each of these models has a distinct set of assumptions that should be considered when developing network models. In the following section, I outline the method I take in investigating the determinant role of exogenous contextual factors.

### 1.4.3 Research Design

The specific network method chosen imposes a number of assumptions on the building of models. Thus in order to accurately provide evidence on the existence of a moderating role for exogenous contextual factors within network structures, a mixed model design is necessary. In answering the question of how context influences networks structures, I take an approach which allows me to investigate effects related to both antecedents and consequences. The use of multiple methods contributing to one shared research question implies the implementation of a sequential mixed-model research design to investigate a number of interlinked propositions. These research designs employ a sequenced combination of models, each of which uses different sets of methods to answer an overarching research question (Ivankova et al., 2006). This allows me to investigate the possible effect of context at different levels of aggregation and implement both static and dynamics models in order to establish:

(1) whether there is a significant interaction effect of contextual factors on the success of individual network structures (Chapter 3);

(2) how specific generative mechanisms are attributed to contextual factors within bounded networks (Chapter 4); and

(3) whether the propensity to interact can be attributed to context at a field level (Chapter 5).

I explain the specific methods employed here below in the section Chapters.

As laid out in the theoretical framework, to explain network structures I consider rules – exogenous contextual factors and a set of resources available to the individual or the endogenous network factors and individual characteristics. Both qualitative and quantitative data are used to measure these concepts (for an overview see Table 1.4). I specify the operationalization of these concepts in the individual chapters and include below a simplified summary of these measures (see Table 1.5). The exogenous context is identified through the identification of polices within an organization to steer a specific organizational behavior observed as a networked relation. The endogenous network factors are identified by network characteristics and individual characteristics identified through individual attributes of the researchers.

Table 1.4: Research Design

| **Data Collection** | **Analysis** | **Presentation** |
|---|---|---|
| Interviews with 17 experts from Dutch Computer Science departmental units | Unpublished data | Qualitatively depict the case study (Chapter 2) and inform specific measurement considerations in the survey (Chapter 3) |
| Interviews with 8 heads of Dutch Computer Science departments | Content analysis | Measure of policies in academic departments and informed survey building (Chapters 2 and 3) |
| Publication data ($N = 3639$ Dutch + Co-author researchers) | Ego network analysis, longitudinal actor oriented analysis, mean field modeling analysis and statistical analysis | Ego network results, longitudinal actor-oriented results, mean field modeling results (Chapters 3–5) and statistical results (Chapters 3–5) |
| Online Survey ($N = 193$) | Statistical analysis | Statistical results (Chapter 3) |
| Collection of bibliometric performance data | Network analysis and statistical analysis | Ego network results and statistical results (Chapter 3) |

Table 1.5: Operationalization of Framework

| **Concept in Structuration** | **Network Theory** | **Measures** |
|---|---|---|
| Rules | Exogenous contextual factors | Policies of an organization |
| Resources | (1) Individual factors,<br>(2) Endogenous network factors | (1) attributes e.g. gender, age, tenure, background<br>(2) network characteristics e.g. previous network structure |

### 1.4.4 Chapters

In this research, to explore the role of context on social network structures, both on emergence and success, I investigate scientific collaboration network structures from multiple perspectives which relate to following empirical chapters. These chapters relate to a number of scientific outputs completed during this PhD project (see Table 1.6). I discuss each of these chapters here as individual methodological steps to answer the research question. Given the shared overarching research question, these empirical chapters contain an acknowledged repetition of descriptions of the theory, method, setting, and data collection.

In Chapter 3, I employ an ego network approach which tests a statistical model of the success of individual scientific collaboration networks while considering a number of policies of academic departments. This static ego network study allows the confirmation of the potential effect of rules on the other resources for an understanding of the network consequences of performance. This provides evidence to confirm an effect of contextual factors on realized scientific collaboration network structures.

In Chapter 4, I seek to investigate how these processes unfold, instead of simply focusing on structural dynamics through the implementation of a SIENA model. As described above, SIENA is the most appropriate technique in considering actor-oriented processes in explaining dynamics (compared to the tie-oriented model of the p*) in addition to its implicit consideration of dynamics in modeling. Although SIENA is not specifically suited for considering exogenous contextual factors in statistically comparing models, I propose a methodological extension. In this extension, I qualitatively compare the results of the individual SIENA models, from which I seek to identify a pattern in departmental policies and the observed generative mechanisms within the scientific collaboration networks of individual researchers. This allows me to use SIENA as a tool to explore the exogenous contextual effect as a moderator for the emergence of different patterns of generative mechanisms.

As discussed above, the Markov chain model used to simulate SIENA has space limitations. An assumption exists that every actor knows the others and has the potential to connect with them, thus implying that the model is often limited by network and model size to accurately simulate network behaviors. Additionally, this limits the statistical comparison for the investigation of the effects of exogenous contextual factors, as the analysis done assumes individuals are within the same bounded context.

In an effort to take advantage of the theoretical assumptions in Markov chain models, I advance the knowledge through a traditionally-used model from the natural sciences - mean field and identify a set of variables that influence dynamics in large social networks in Chapter 5. Using models from Chapter 4, possible conditions that influence dynamics are aggregated into a larger model, thus averaging effects to identify main influencing dynamics. As a result, the context here is no longer a control or boundary condition, but an explanatory factor in explaining dynamics. This process effectively advances a global perspective on identifying conditions, isolating only those variables that influence a majority of individuals. Findings show that the department does play a role in collaboration with other departments, confirming results from previous chapters: an exogenous contextual factor plays a key role in the *modus operandi* of social network dynamics.

The culmination of the findings from the three empirical chapters is reflected upon in Chapter 6. All three of these empirical chapters are explored through the lens of Dutch Computer Science researchers scientific collaboration patterns, explained here below and in further detail in Chapter 2.

### 1.4.5 Setting

Science rests on the perpetual quest for knowledge and requires observing, exchanging, discussing, and reflecting on knowledge (Knorr-Cetina, 1981), as well as access to resources (such as codified, tacit information), which help to develop new knowledge (Latour and Woolgar, 2013; Polanyi, 1967). Knowledge production is pursued through the lenses of disciplines. These disciplines regulate conduct, which aids in organizing the practices and boundaries of specialties (Foucault, 1977). In turn, disciplines are sustained through formal communication practices such as the dissemination of knowledge to field-specific journals, conferences, and reports to scientific peers. These peers are an "invisible college" of actors where an audience receives knowledge and reacts to it. They are "invisible" in the sense that actors producing the knowledge may not physically interact with or know of one another but contribute nonetheless to the same field of knowledge. They, thus, are connected through citations and references to similar works (Crane, 1969). Academic institutions, organized around these disciplines, serve as cornerstones for organizing education and work practices. They provide an environment for the pursuit of knowledge production. Within these formal academic institutions, the practices of knowledge production are further classified into faculties, departments, research institutes and research groups. They are the localized seats where researchers pursue scholarship. Consequently, the practice of science is social; it is arguably related to individual researchers navigating their actions between an ensemble of exogenous and endogenous factors for the pursuit and production of knowledge. These patterned practices of work, maintained through social construction, application and reiteration, create and maintain disciplinary boundaries while maintaining the practices of research (Whitley, 2000).

Recent decades have seen the emergence of new modes of scientific organization within all disciplines of scientific practice (Gibbons et al., 1994). These new modes are evident from the increasing prevalence of co-authorship on academic publications (Greene, 2007), more teamwork in science (Falk-Krzesinski et al., 2010), growing cross-sector cooperation (e.g., triple-helix configurations (Leydesdorff and Etzkowitz, 1996)) and shared laboratory pursuits (Shrum et al., 2007). The increase in collaboration in science has been attributed to its perceived positive benefits, such as those in the division of labor (Hudson, 1996) that yield improved ”market value” for scientists and their work (Sauer, 1988). Furthermore, the increased ease of connectivity via online communication and increased travel of researchers play important roles in such collaboration (Ding, 2011). These new modes of scientific organization have implications for the nature and quality of knowledge production (Gibbons et al., 1994). Thus, scientific organizations, academic institutions and policymakers are seeking ways to facilitate and encourage collaboration through teams and formal research initiatives such as grants and institutes (Defazio et al., 2009; Guimera et al., 2005).

Consequently, a large body of research has investigated scientific collaboration. The manner in which researchers employ different network structures to develop publications has most often been studied as an individual-level phenomenon, in which researchers’ collaboration tendencies and performance are explained from their formal position, gender, and experience (Bozeman and Corley, 2004; Lin and Bozeman, 2006), although location (e.g. the academic department) retains a strong positive effect on the performance of researchers (Long, 1978). Recent studies suggest that external factors influence the structures of and tendencies towards collaboration (Gibbons et al., 1994), such as where institutional factors (Boardman and Corley, 2008; Ponomariov, 2007; Shrum et al., 2007) and the requirements of funding schemes (Defazio et al., 2009) serve as moderators for the success of the network performance of individuals. Thus there is a question as to the role of contextual and individual factors in collaboration networks, making it an ideal match with the gaps investigated in this research.

#### 1.4.5.1 Scientific Collaboration & the case of Dutch Computer Science

The use of scientific collaboration to explore these social network dynamics has a number of advantages. At a basic level, science provides a number of clear organizational entities that can be considered potential rule makers in measuring exogenous contextual factors. These entities could include: the practices of the discipline, (supra) national science policies and regulations, institutional policies, and departmental polices. These bodies implement policies and regulations in an attempt to direct researcher behavior. I will not discuss the nature in which such policies come to fruition within these organizational bodies. Instead, I focus on the network structures that emerge in light of these policies.

At a next level of consideration, we see how these organizational bodies provide clear boundaries in defining exogenous contextual effects, which, in turn, allows for a valid and reliable method for selecting a context. Take, for example, the policies established by the department in which a researcher conducts his research. These policies are the criteria for ”what it takes to be part of the club” where researchers are evaluated, measured and/or reviewed on the quality of their work through tenure systems, target lists, teaching evaluations, and so forth. These policies weed out researchers of a particular type who do not meet the criteria designated by this system. Consequently, conceptualizing the exogenous contextual factor of a department instituting such policies as a determinant to the emergence of a specific scientific collaboration allows me to investigate if and how these specific conditions lead to researchers employing different collaboration network structures. I explain in detail in Chapter 2 and in the individual empirical chapters pending the variable measured the policies used in Dutch Computer Science departments in particular as instruments to alter the production of knowledge of researchers.

Scientific collaboration also provides a valid setting in which to explore these dynamics. First it is seen as a phenomenon that is driven by the relationships that individual researchers establish in their pursuit of knowledge. In this regard, scientific collaboration is most often measured via co-authorship (Melin and Persson, 1996). Publications acknowledge a formal collaboration that infers interactions to realize new knowledge. Publication and bibliometric data provide a valid and reliable source for measuring both networks and network changes, as well as outcomes such as the performance of these co-authorships in a particular publication. The nature in which this knowledge is appreciated in the field is via citations in other publications. This provides a valid marker of performance by which individual researchers are evaluated. In addition, a large amount of descriptive information is publicly available about researchers (e.g., affiliations, career, age, and gender).

In this research, therefore, the setting of science, as described above, provides the background for our case study. Such a case provides a lens for both examining concepts and generalizing to other fields (Thomas, 2010). The use of one case provides a number of clear boundaries, governing rules and resources. I sought a case that would not only be generalizable to social network phenomena, but would provide clear boundaries for the identification of factors within the context, as well as within a set of individuals and networks. One field was selected, as well as one national context, therefore, in order to limit variance in the exogenous contextual factors affecting scientists.

For these reasons, I investigate individual scientific collaborations of Dutch Computer Science researchers. The Netherlands was selected because it is a typical European academic environment with funding on both national and international levels for the purpose of stimulating research. Such an environment also provides a diversity of cases for examining different institutional processes in a relatively small geographical space. In addition, the Dutch context was selected to take advantage of the opportunities provided by my own in-depth local knowledge, given the location of the PhD project and my own research interest in cross-disciplinary work with the field of computer science. Within the Netherlands, computer science is a field of high-quality research, with significant though decreasing funding allocated for disciplinary-focused work (Nationale Informaticakamer, 2010). Additionally, the field of computer science was chosen for a number of other more historical and logistical reasons:

(1) the general maturity of the field having a number of distinct sub-fields;

(2) its known tendency for collaboration through co-authorship;

(3) the validity and reliability of online sources documenting publications;

(4) my own personal familiarity with the Dutch science system and thus access to researchers to conduct the research.

Computer science is a field based on both information and computation studies, both of which join forces in the use of computers as systems or tools for solving research problems. The discipline of computer science is a mature, intellectually unified field with a number of mature sub-fields existing as self-sustaining practices. Computer science subjects range from theoretical computer science and information systems to software engineering and computer networks. Consequently, the field not only works on internal questions but also has a tendency to work outside its realm.

Within the case of Dutch Computer Science, I select a number of academic departments. The computer science departments at academic universities are the smallest units of an exogenous contextual form that would potentially have policies dictating individual behaviors, particularly those related to scientific collaboration through publication. This provides a number of different combinations of contexts in which to explore the effects of exogenous contextual factors on the individual scientific collaborative behaviors of researchers. I elaborate on this case and the precise sample of individual researchers in Chapter 2.

## 1.5 Conclusion

This introduction outlines the approach taken in the investigation of the effects of exogenous contextual factors on social network structures through the lens of scientific collaboration networks of individual Dutch Computer Science researchers. In reviewing theory on networks, questions were advanced based upon the roots of the mechanisms. A focus on individuals as drivers of structures in network theory has been a detriment to the study of context, leading to a blind spot in current knowledge on network structure emergence and success. A framework for such a study integrating structuration and network theory was stipulated as to how context – as understood as rules – and individual and endogenous factors – as resources – interact to influence the structure undertaken by individuals to realize action. In this framework, context serves as a detriment to the emergence of network structures. A methodology was discussed where three techniques were reviewed. A design was developed based on mixed methods, allowing me to take advantage of the different ontological assumptions of current methods in exploring the effects of exogenous contextual factors, investigated in three empirical chapters.

These findings contribute to the theoretical understanding of the antecedents and consequences of social network structures, providing insight into how to understand network structures in the consideration of context as a determinant. In view of how the individual perspective in collaboration has been the focus in recent studies in science, there is also a need for a study focusing on the policies and boundaries in the practice of computer science in The Netherlands. I reflect on these findings in the conclusion of each chapter and further in the concluding chapter. This research contributes to social network theory and method, as well as practical knowledge about the emergence and success of scientific collaboration networks in describing the *modi operani* of social networks dynamics.

Table 1.6: Scientific Outputs Related to Empirical Chapters

| **Chapter** | **Related Outputs** |
|---|---|
| Chapter 2: Dutch Computer Science: A case for investigating dynamics of scientific collaboration networks | This chapter is part of a working paper on Dutch Computer Science prepared in cooperation with the Rathenau Institute, The Netherlands. |
| Chapter 3: Context, Network, & Performance: Contingencies of Successful Collaboration Networks | This chapter is a working paper and a candidate for publication consideration, and is related to work presented by:<br>1) Birkholz, J.M., P. Groenewegen and E. Horlings (2014). Collaboration's many forms: context, networks and performance. Presentation at the Workshop on Research Funding and the Dynamics of Science, by Research Network 24 – Sociology of Science and Technology Network (SSTNET) of the European Sociological Association (ESA) and the Centre for Science and Technology Studies (CWTS) at Leiden University, The Netherlands. |
| Chapter 4: Considering context: recasting the SIENA model to consider contextual factors as determinants of social network structures | This chapter is a working paper and a candidate for publication consideration, and is related to the following:<br>1) Birkholz, J.M., P. Groenewegen and P. Groth (2012). The Role of Status in Team Formation in Science. Poster at the NetSci Conference, Northwestern University, Evanston, IL, USA; and<br>2) Birkholz, J.M., M. de Klepper, P. Groenewegen and P. Groth (2011). Dynamics of scientific collaboration networks. Presentation at Sunbelt Social Network Analysis Conference, Palm Springs, FL, USA. |
| Chapter 5: Scalable Analysis for Large Social Networks: The Data-Aware Mean-Field Approach | This chapter was published in its current form (Birkholz et al., 2012), and is related to research presented by:<br>1) Birkholz, J.M. (2012). Scalable Analysis for Large Social Networks. Invited speaker as part of the SONIC speaker series, Northwestern University, Evanston, IL, USA; and<br>2) Birkholz, J.M., A. Lungeanu, R. Bakhshi, P. Groenewegen, M. van Steen and N. Contractor (2013). Methodological Specifications for Application of the Mean-field Model for Large Scale Social Networks. Presentation at Sunbelt Social Network Analysis Conference, University of Hamburg, Hamburg, Germany. |

# Dutch Computer Science: A case for investigating dynamics of scientific collaboration networks

2

In this dissertation I investigate the effect of context on the network dynamics of individuals within the setting of academic science. Science provides a valid and reliable setting for defining possible formal organizational boundaries at which to isolate factors of the exogenous context, as well as to clearly identify a set of research individuals level whose networking behaviors can be observed. As explained in detail in the Introduction, I implement a case study to analyze the dynamics of scientific collaboration networks by investigating Dutch Computer Science researchers.

This case study is used to exemplify a social environment suitable for the exploration of social network dynamics, not as a methodology to directly comment on the field of Dutch Computer Science. This is not to serve as an evaluation nor a bibliometric review (for such reviews see (Bar-Ilan, 2010a; Franceschet, 2010; Goodrum et al., 2001)). Nor is this case description used to theorize specifically about science dynamics or science theory; I leave theorizing of this nature to work related to (Kuhn, 1996; Lakatos and Musgrave, 1970; Toulmin, 1972). In order to situate these findings in the field I present here a general description of Dutch Computer Science.

## 2.1 Sample

The field of computer science was selected for three reasons: the tradition of diversity of sub-fields within the discipline, the known tendency for collaboration through co-authorship and the validity and reliability of online sources documenting publications. Computer science is a field based on both information and computation studies; these converge in the use of computers as systems and/or tools for solving research problems. computer science as a discipline arose in the 1940s with the emergence of mathematical logic and the invention of the "stored-program electronic computer" (Denning, 2000). As Denning (2000) described, the discipline centers around the question of "What can be (efficiently) automated?". Thus, in the broadest sense researchers work on establishing methods or procedures for accessing, storing, processing, and representing information, although the field is methodologically, ontologically, and epistemologically diverse (Eden, 2007; Wegner, 1976). The tasks that a computer science researcher undertakes have a large span with diverse data types (e.g. bits of memory, communication networks, Semantic Web data) and different types of automation processes (e.g. algorithm development and improvement, mechanization of processes), as well as a diverse set of problem-solving tools (e.g. hardware, software, languages). Thus, researchers adopt different kinds of conceptual, theoretical, methodological, and philosophical frameworks for each research study (Tedre, 2006). (For further insights into defining computer science see Newell and Simon (1976), and for a review of the development of a philosophy of computer science, see (Tedre, 2006).) The discipline of computer science is seen as a mature, intellectually unified field with a number of mature sub-fields existing as self-sustaining practices. This range of sub-fields in the discipline suggests that the those within the discipline not only work on internal questions but also have a tendency to work with other fields, increasingly in teams (Wuchty et al., 2007).

In order to establish explicit boundary conditions for studying dynamics, I selected one nation's endeavors in the field, namely Dutch Computer Science researchers, focusing on researchers from nine academic computer science departments in the Netherlands; these provide the boundaries for the definition of contextual factors. The Dutch context was selected for it is a typical European academic environment with funding at the national level for research that stimulates cooperation; it provides a diversity of cases which allow for the examination of different institutional processes in a relatively small geographical area. Dutch Computer Science departments are high quality with many instances of excellence (Nationale Informaticakamer, 2010). In addition the Netherlands was selected as a case to take advantage of the opportunities provided by my own in-depth local knowledge. It should be noted that I am informally affiliated with the Department of Computer Science at the Vrije Universiteit Amsterdam via the professorship of promoter Prof. dr. Hans Akkermans. That being said, given that I was interested in identifying the rules in these contexts and not evaluating them, my affiliation provided few subjectivities and rather a perceived advantage. This advantage allowed me access as a peer, instead of an outsider. Additionally, my close knowledge of the field from first-hand experience provided an advantage in understanding field-specific terminology.

To increase validity of the units selection, I used the same selection as the *Nationale Informaticakamer*'s five year evaluation of Dutch academic research institutions, choosing the following nine academic research universities:

- Technische Universiteit Delft,
- Vrije Universiteit,
- Technische Universiteit Eindhoven,
- Universiteit van Twente,
- Universiteit Leiden,
- Radboud Nijmegen Universiteit,
- Universiteit van Amsterdam,
- Universiteit Utrecht,
- Rijksuniversiteit Groningen.

Within these universities computer science is formally organized in natural science faculties or institutes and then further organized into smaller thematic units as departments, sub-departments, chairs or groups. It should be noted that in two universities two separate units working on computer science exist which share faculty affiliation. Given the level of abstraction required in this research (classifying contextual factors), classifications at the level of a shared faculty of individuals pursuing computer science serve the research purpose and mimics distinctions made in formal evaluation reports (Nationale Informaticakamer, 2010). For the complete list of departmental units as of 2012 see Appendix Table A.1. Within the empirical chapters these departments are anonymized to protect the researchers.

Given the period for the dissertation research the identification of departmental units via *Informaticakamer* used 2010 as middle point of the observation period and data collection. As outlined in the Introduction I compiled data from multiple sources, some of which rely on the reporting of individual researchers through interviews and surveys. Additionally, in order to increase validity I limited the reflection period to 5 years, focusing on scientific collaborations presented in publication data of the 2006 - 2010 period. In some chapters, data from the 2006 - 2012 time period was reviewed to account for publication and citation lag (see the methods sections of each chapter for the specifications). This choice coincides with national funding changes that occurred in 2005 and 2011, described below in detail under ”funding”. Accordingly, this study indirectly characterizes a specific era of Dutch Computer Science.

I initially identified a group of individual researchers within these departments from a list of known active researchers compiled from two sources:

(1) tenured staffed researchers in 2010 provided by Narcis (NARCIS, 2014) (again coinciding with the year of the *Informaticakamer* review), and

(2) directories of all staff working at the above-mentioned nine computer science departments as listed on the affiliated academic institutional websites (as posted September 2012).

From this list, publication data was queried using *Digital Bibliography & Library Project Database* (DBLP) (Database Systems and Logic Programming, 2013). DBLP is one of the most comprehensive bibliographic indices for the field of computer science and thus allows the identification of a valid and reliable set of co-authorship publication data for Dutch Computer Science researchers. This query resulted in a list of both Dutch researchers and their co-authors. In order to confirm researcher affiliation and increase the validity of the resulting list, the institution was identified through queries of two additional databases. The automatic collection of historical data on institutional affiliation is not currently stored in one database. A query using Microsoft Academic Search – a database which includes the DBLP data set – was used to identify institutions (Microsoft Academic Search, 2014). To locate additional missing data, ArnetMiner.org (ArnetMiner, 2012) was used. This database is a search and mining service of computer science researchers which includes semantic data on computer scientists. In order to disambiguate institutional names and have a reliable and valid set of data, the resulting list was further queried in the geocoding Web service Yahoo! PlaceFinder (Yahoo! PlaceFinder, 2012). This geocoding Web service converts street addresses or place names into geographic coordinates. This query provided a proximity measure for each institution and a uniform institutional affiliation based on common GPS coordinates. The final result of these queries led to the identification of 1516 Dutch Computer Science researchers. Each of the three empirical chapters of this research reflects the use of a sample of this source data and publications for collecting additional relevant data in relation to the research question (the measurement of these variables are discussed in detail in each of the empirical chapters, given the research question).

As discussed in the methods section of 1 I measured a number of conditions of these units in regards to the research question. Identification, us and exact measures of concepts are specified in the individual empirical chapters.

Thus, the case study undertaken describes the field of computer science through the analysis of individual computer science researchers' scientific collaboration networks in The Netherlands. A case study that represents one field, one nation, and a specific and identifiable set of individuals provides increased validity for specify dynamics and the exploration of the research question. It also provides added validity in the implementation of a mixed method model research design that, in particular compares findings from empirical studies. The use of one case allows me to expand on the largely scientific findings of the dissertation research and also contribute to broad policy implications for Dutch Computer Science.

## 2.2 Method

The depiction of Dutch Computer Science is based on both findings from unpublished interviews and a number of tertiary sources on computer science – Dutch Computer Science in particular. The findings in this review reflect the nine academic research universities investigated in this research.

Qualitative interviews provide a tool for undertaking exploratory research and allows the researcher to develop detailed descriptions of conditions, integrate multiple perspectives and thus bridge inter-subjectivities from multiple parties into a coherent story and describe a process, as well as develop holistic descriptions (Weiss, 1995). Interviewees were identified through purposive sampling to select two experts in each of the nine universities; cross checking provided a selection of two computer science researchers from two different fields who together comprise more than 50 years of work exclusively within The Netherlands. Expert research staff members are defined as holding an associate professorship or higher with 5 or more years of research experience in the department. These criteria insured multiple perspectives on the field, institution and the department. In total, 18 experts were identified, with 17 experts agreeing to be interviewed; one department head declined to be interviewed. Sixteen interviewees had head professor status and one was associate professor. The interviews were conducted between Fall 2011 and Spring 2012 and took place either in the offices or off campus place of the interviewees' choosing. The interviews lasted on average one hour each and were conducted solely myself. The majority of interviews were recorded, and notes were always made during the interviews. In a few cases the interviewee asked that the recording be stopped, largely during our discussions of policies and/or funding, but during those instances notes were taken. These interviews were anonymous and as promised to interviewees, following a summary of the reports I will make sure to destroy recorded data so as to not have their voices recognized and attributed to the research.

The goal of the interviews was three-fold:

(1) provide an overview of the field of computer science in The Netherlands,

(2) provide insights into how scientific collaboration was occurring, and

(3) identify the policies on scientific collaboration through publication that impacted researchers within the department.

The interviews were semi-structured, not only to collect standard data on policies but to also allow interviewees the freedom to tell their stories. This allowed me to get a perspective of what was going on in the field presently, as well as collect variables on the independent variable of exogenous context as described in Chapter 3 and 4. I summarized every interview and analyzed the content of the interview data to find common themes and events mentioned that explain key occurrences within field. The tertiary sources were sourced following these findings. For example, if an interviewee mentioned that changes within a specific funding body altered the field in some way, I then later sought official documentation to provide further context on the specific funding body in order to understand the field in general. Thus the data is presented here as a historical summary, not as direct quotes. They are summaries that reflect a shared review of Dutch Computer Science from 2006 - 2012.

## 2.3 Dutch Computer Science

When asking Dutch Computer Science researchers about the history of Dutch Computer Science, a majority of them mentioned Edsger Dijkstra, winner of the prestigious Turing Award, mathematician and later self-proclaimed information specialist who played a key role in the emergence of computer science, *Informatica*, in The Netherlands. It was not only his academic work that put Dutch Computer Science in the limelight but also his outspoken conservativeness about the directions of computer science, Dutch technical investments and academic education. Ironically, despite being mentioned as a type of father for Dutch Computer Science, Dijkstra left Dutch Computer Science in the early 1980s for the University of Austin, Texas, USA, at the same time that computer science was beginning to be recognized by many universities, both in The Netherlands and abroad.

In the late 1970s and early 1980s, a number of departments were founded.[f] Before this time those interested in computer science were *hidden* in departments of mathematics, physics, and engineering where they were often given special chairs to "deal with the computer things". In fact a large portion, arguably the largest, of tenured computer science researchers do not have backgrounds in computer sciences, as we know the field today, but rather were trained in other natural sciences and had a growing interests in automating solutions. Most academic units started working on topics of theoretical computer science, which evolved into (in no particular order) work in information systems, software engineering and computer networks.

Coinciding with the emergence of the research units was the responsibility of educating those in the emerging field. The first formal computer science educational programs at the Vrije Universiteit Amsterdam and the Universiteit van Amsterdam began in 1981. Presently, all of the computer science units in this case study provide both Bachelor's and Master's degree programs. These range from general *Informatica* degrees and degrees in other natural sciences, as well as minor specializations such as security or artificial intelligence. Researchers are largely responsible for both education and research. Rising interest in education programs resulted in an increase in political power for computer science researchers in The Netherlands within their universities, in general, and within fac-

[f] During this period the legal term for "departments" was *vakgroepen*, which are now smaller units with less formal distinction.

ulties in particular. Since the majority of universities assign a proportion of staff members to a unit based on the number of students and/or number of courses, this is the *first stream* of funding. Thus the more interest and students, the more staff. Over the last decade, this schema has shifted in two distinct ways:

(1) the emergence of separate teaching groups where researchers have 100% teaching obligations, and/or

(2) researchers being able to "buy off" teaching time upon the receipt of a grant from a different *stream*.

As society's interest in technology grew and investment in technology from both private and public sectors increased (discussed below specifically under Funding), so did the field of Dutch Computer Science and the acknowledged importance of computer science throughout the world. This helped to support the emergence of a number of specializations in The Netherlands. These specializations include (non-exclusive list, in no particular order): artificial intelligence, bioinformatics, human and computer interaction, multimedia, robotics, ergonomics, information systems, management information systems, business applications, enterprise computing, embedded systems, hardware and technical engineering units (e.g chip development), and digital security. The three technical universities largely have a monopoly on hardware and engineering domains, but there remain specializations within the universities, particular those combining expertise, for example, in the specialization of robotics.

## 2.4 Publication Practices

The diversity of specializations in computer science, and Dutch Computer Science in particular, lead to an acknowledged often perceived fragmentation of the field, particularly by outsiders. Interviewees acknowledged that describing to outsiders how computer science researchers were not like plumbers is a problem of the field in general, not just Dutch Computer Science. This discrepancy was largely attributed to the lack of recognition of the distinct communication practices of computer science.

Unlike other natural science fields, such as the life sciences and physics where peer-reviewed journals make up the majority of dissemination practices, conference proceedings comprise the largest share of publications within computer science (Franceschet, 2010; Goodrum et al., 2001; Moed and Visser, 2007). In the fast-advancing fields of technical sciences, conference proceedings are a legitimate and accepted form of communication that serve as an end stop for knowledge, and fill an important gap in academic knowledge production (Drott, 1995). This is particularly so because conferences proceedings, like journal publications, undergo a peer-review process and lead to the acceptance or rejection of publications (but usually at a more-timely rate so that ideas can quickly be claimed with documentation). These proceedings publications are also among the highest cited publications within computer science (Moed and Visser, 2007). The largest and most commonly used bibliometric indexes, such as Web of Science (WoS), only cover a small number of proceedings publications. Among those proceedings are: Lecture Notes in Computer Science (LNCS), which includes a number of computer science proceedings – but recommendations have been made to overcome this, such as an expansion of Thomson Reuters' Web of Science database to include proceedings coverage as suggested by experts, as well as adding bibliometric data about proceedings and books from the digital libraries of the Association for Computing Machinery (ACM) –, and the Institute of Electrical and Electronics Engineers (IEEE) (Moed and Visser, 2007). In 2008 WoS added the Conference Proceedings Citation Index-Science, with proceedings from scientific and technical fields, and Conference Proceedings Citation Index-Social Science & Humanities, with proceedings from social sciences, arts, and humanities, both covering proceedings from 1990 to present (Bar-Ilan, 2010a). This addition, saw an average publication count increase of 39 percent for the field of computer science, but a large variations between researchers given their sub-fields was acknowledged (Bar-Ilan, 2010a). Despite the known value for the knowledge production system of computer science, proceedings are often absent from commonly used indexes and formal reviews and evaluations by faculties and universities. This distinct communication practice is often misinterpreted as a way to publish low quality or findings of less theoretical or practical value.[g]

As evaluations are largely conducted measuring citation scores of a standard set of journals, this lack of recognition of conference proceedings often leads to the lower evaluation of computer science researchers when compared to their peers in the natural sciences. This presents a misleading perception of the value of the knowledge produced by computer science as a field. Interviewees acknowledged a constant dialogue with deans in attempts to explain the value of proceedings in the field and the, thus, discriminatory nature of current metrics in comparing computer science to other natural sciences, and the employment of policy decisions such as funding allocation and restructuring. Consequently, the lack of consideration of conference proceeding publications remains an issue for the perceived legitimacy of Dutch Computer Science researchers.

## 2.5 Embedding

When asked during the interviews about collaboration in computer science as a broad topic, a number of different subjects emerged. I consider this the formal and informal embedding of researchers within different types of cooperation and affiliations as it did not necessarily related to the scientific collaboration via a shared publication investigated in this research. It is not my aim to mention here

[g]Parts of this paragraph are drawn from an unpublished paper: J.M. Birkholz & P. van den Besselaar (2012) "Comparing apples and oranges: Using informed bibliometric indexes in the case of Computer Science", where I was the contributing author.

an exclusive list of those with whom a unit formally cooperates, but rather to provide examples as to the type of affiliations and organizations that exist within Dutch Computer Science.

The organization of the departmental unit varies by university. In addition to the differentiation through the titles of these units as presented in the Appendix, Table A.1 (e.g., some are recognized as institutes while others are departments), the roles of these units different. I do not intend to identify every distinct organizational difference between these units but rather to acknowledge that research and education is organized in different ways among the universities. Largely in most cases the unit as the department is responsible for both education and research and reports with a management team to a dean of a faculty. Increasingly, within the universities there is also a rise in the emergence of inter and cross disciplinary research institutes in an attempt to stimulate research between disciplines. These institutes have varying reach and formalization, with some steering research agendas with required affiliation and others more informal with their own budgets for activities to stimulate research. In some cases, for example at the Universiteit van Twente, researchers are affiliated formally with both the department and an institute (e.g. Center for Information and Communication Technology – CTIT (CTIT, 2014)). The department is responsible for education and the institute is responsible for research, both reporting jointly to a dean of a faculty. In other cases, for example at the Vrije Universiteit Amsterdam, research institutes (e.g. The Network Institute (2014)) serve as a body to organize research between faculties, and are exterior to research on the guise of faculties. The perceived effect of these institutes varied by interviewee, allowing me to suggest that it had less to do with the institutes themselves but rather the orientation of the research(er).

On the national level, individual researchers are affiliated via the research schools. These research schools emerged in the 1990s to support PhD training and facilitate the development of national research through the building of informal bridges among researchers throughout the country. The three largest research schools are the Advanced School for Computing and Imaging (ASCI) (ASCI, 2014), the Institute for Programming Research and Algorithmics (IPA) (IPA, 2014) and the School for Information and Knowledge Systems (SIKS) (SIKS, 2014). A large portion of researchers are affiliated with teaching courses and providing general support for the school by way of directing research lines through the investment of junior researchers.

In addition, the nine departmental units have various affiliations and formal cooperation with other departments, faculties, universities and countries through a number of thematic networks. For example the association of three technical schools within the Netherlands, 3TU (2014) aims to serve as a contact point for facilitating the exchange of research topics among technical schools. A similar federation exists for technical school at the European level, where lobbying at EU bodies also occurs. Additional European initiatives exist with field-focused associations such as Informatics Europe (Informatics Europe, 2014). Interviewees questioned the larger impact of these bodies in shaping research, policy or education, but they were generally perceived as formal contact points that were often activated when in need of partners.

Thus researchers have a number of ways to embed themselves in conducting their work. These are also potential sites for interaction with potential scientific collaborators.

## 2.6 Funding[h]

The nature in which academia is funded in The Netherlands also plays a role in understanding the case of Dutch Computer Science. The Dutch government funds 14 universities that are responsible for both scientific research and education (Rathenau Institute, 2014b). Academia is funded via three *streams*: the first, second, and third funding schemes (*eerste, tweede en derde geldstromen*). In this section I do not aim to review the entire science system of The Netherlands nor speculate on the effects of funding individual universities, but rather review the main funding sources and changes specifically for computer science within this system.

The first stream of funding comes directly from the Ministry of Education, Culture and Science (*het ministerie van Onderwijs, Cultuur en Wetenschap (OCW)*). This is a sum based on the number of students and a number of parameters for research. At universities, funding is largely allocated based upon the education responsibilities of departments, and, despite attempts at transparency, the proportion allocated is also largely related to political processes. Within departments this funding is principally allocated for permanent positions such as head professors and support staff. Thus, again there is an emphasis on the continued concern for the status of computer Science to receive continual funding to develop a sustainable research line.

In addition to this funding, the government also funds the second stream of funding via the National Science Organization (*Nederlandse Wetenschaplijke Organisatie (NWO)*) and the Royal Dutch Academy of Sciences (*Koninklijke Nederlandse Akademie van Wetenschappers (KNAW)*). The KNAW is largely responsible for supporting head professors. The NWO receives the largest amount of funding, which it has the task of redistributing in the form of grants for individual researchers and research teams based on a number of scientific themes; consequently it is also where the majority of researchers (traditionally) receive funding.

These themes relate to domains with funding mechanisms for specific investing in the enhancement of specific knowledge infrastructures. Arguably the largest investment in computer science in this second stream funding emerged between the early 1990s and 2010 via research domains that fell under invest-

[h]The information about funding streams in this section is largely sourced from general information about the Dutch academic sector from the Rathenau Institute. See: (Rathenau Institute, 2014b).

ment of *het Fonds Economische Structuurversterking (FES)* (Grants for Improvements of Economic Infrastructure) and later *het Besluit Subsidies Investeringen Kennisinfrastructuur (Bsik)* (The Directive for the investment in knowledge infrastructure subsidies). FES funding was established in 1994 from the profits of natural gas exploitation as an effort to stimulate Dutch knowledge infrastructure. Two rounds of grant funding occurred in 1994-1998 and 1999-2003 and again under Bsik from 2004 to 2010. Many of these projects required co-funding from either the universities or private companies to stimulate cooperation. This funding was terminated in the government coalition agreement (*het Regeerakkoord*) of October 2010; thus only a few programs remain funded from this mechanism at present. Within the Bsik program 215 million euros went specifically to projects in the technical sector. (Rathenau Institute, 2014a)

Many interviewees acknowledged a shift in this tradition of NWO funding for the field of computer science following changes in 2005. At that time, the NWO restructured thematic funding, combining computer science and astronomy. This put computer science in competition with so-called fundamental research. Many found this put new demands on Dutch Computer Science researchers to position themselves within a theoretically and methodological distinct field such as astronomy. Thus Dutch Computer Science researchers again found themselves compelled to define their field and domain application in relation to a fundamental science field such as Astronomy.

In addition to NWO as a main redistributor of academic funding, a number of tertiary organizations exist for supporting research in computer science. These organizations are funded within large government investment programs (e.g. *NWO Grootinvestment programme*). The government aligns its goals with restricted funding proposals, ones which require the management and organization of preferred organizations. These include the Technology Foundation STW and COMMIT. These foundations (*stitchingen*) fund, manage and organize ICT research, education and cooperation initiatives in The Netherlands.

The third stream of funding is financing from other public or private sources. Dutch Computer Science researchers are increasingly funded from the European Union. Funding in the 1990s for Dutch Computer Science was largely from the European Strategic Program on Research in Information Technology (ESPRIT), which funded a number of technology investments (European Communities, 2014). Most Dutch researchers did not seriously become active in pursuing European Union funding until the 6th and 7th Framework, which also specifically included technical investments. These mechanisms largely supported research in consortium with multiple European universities working on both soft and hard infrastructure projects. Interviewees acknowledge EU funding as a stable source of increasing funding for the field, although it has presented a number of new challenges when compared to NWO funding as EU funding often demanded extensive grant documentation, identification of a consortia and project management.

Third stream funding also includes funding through cooperation with the private sector. Increasingly, researcher positions were supported by private companies where researchers physically worked a few days each in both a company and at the university. In addition private companies often supported specific project-based initiatives or supported PhD candidates through internship types of access to data, equipment and the like. These types of grants were most common at the three technical universities.

These processes in acquiring funding are not distinct to this case, but certainly the issues of identity and recognition add an additional layer to the complexities of this process. This issues were stipulated in the Dutch Government's Innovation Policy of 2010 (*top sectoren*). The Innovation Policy of the Rutte I cabinet in 2010 named five strategic sectors where investments would specifically be made. This policy largely supported projects between the private and public sector. It received criticism given its focus as a "coordination instrument" instead a funding mechanism as done in past innovation funding thus also changing the application scheme where researchers, and increasingly non-junior researchers, applied in consortiums (van den Toren et al., 2012). More importantly for this case, there is no specific mention of the role of computer science in innovation in The Netherlands. Many researchers saw this as a blow to the perceived role of computer science in society and the perceived importance it had among the government in stimulating innovation. Additionally, it was seen as a potential threat to sustainable national funding for computer science given the unspecified role of ICT in the five strategic sectors of investment, as funding for innovation falls only in these sectors. In response to these concerns and as a method to define that role, the ICT-Roadmap emerged in 2012. This initiative sought to highlight the implicit role of ICT in innovation and the specific links needed between academic computer science and technical companies in the private sector for the achievement of the goals stipulated for the five sectors (Lundqvist et al., 2012).

Due to changes in the structure of funding, particularly in first and second stream funding, researchers in general, but particularly in The Netherlands, have increasingly less job security (Kuiper, 2014). The increase in project funding (and thus the decrease in first stream funding for academic institutions) influenced the research environment in two ways. First, the decrease in funding allocated for discipline-specific, focused work is a potential threat to the health of the field in The Netherlands (Nationale Informaticakamer, 2010). Secondly, interviewees acknowledged that this leads to contentious political decision-making in the consideration of hiring research staff for permanent positions, as funding for these positions have become more and more scarce; additionally, a permanent chair guarantees a thematic permanency. In a response to these uncertainties in many cases researchers find autonomy through sourcing their own research grants from various bodies and buying off their required teaching time (This is discussed in further detail in the following section concerning professional tenure systems.) Consequently, the acquisition of personal grants provides temporary certainty in the development of a research line, but not necessarily a track to a permanent position.

## 2.7 Policies

As outlined in the Introduction, in identifying contextual factors I sought to identify policies that serve as rules for guiding scientific collaboration behavior. The university, and more specifically the faculty and departmental units, have a number of instruments that steer researchers outputs in particular directions. Many researchers thought that scientists themselves known the field better than the management systems that attempt to steer them. Three key *rules* emerged from these findings:

(1) a professional tenure system,

(2) preferred publication outlets, and

(3) incentives.

The makeup of these policies differed between departmental units. For the purposes of describing the case of Dutch Computer Science, I describe these policies in general terms here; in the chapters that follow they are discussed in more detail.

First, a professional tenure system outlines the specific core tasks of their employees and how they will be considered. Some of these policies are contractual. These formal conditions stipulate the process through which an employee is evaluated and the process of promotion, firing and hiring. Such policies do not guarantee a position but instead mandate a due process and a way to obtain a specific status. For example, in teaching and academic professions, a professional is on a tenure track, similar to the conditions for a lawyer to make partner. Such a professional system aids in defining who is in and out of the club, given they follow a set of rules.

Tenure systems for computer scientists have emerged in the last five years, largely in Dutch Computer Science faculties. Some departmental units do not have professional tenure systems, and the levels of formalization, specification and length of existence vary. In academic science, professional systems of this nature set a standard of research quality and attempt to steer the output of researchers. The system stipulates guidelines for promotion, and is largely a top-down process in universities or faculties. In practice, this system is much more tedious than a simple “rulebook”. As described above in the funding section, promotion is largely political given increasing financial uncertainties. Thus there is an acknowledged “generation gap” between two levels of staff, where those already in the first stream never went through such political tedious professional systems nor had to prove themselves largely from publications and grants, as do those in the later streams. Only in a few cases where a tenure system already existed were tenured staff subject to these systems; others were grandfathered in. Few public documents were available as to what stipulates professional tenure conditions. I attribute the sensitivities around the procedures of tenure to the political nature of social advancement. Achieving tenure is a tenuous process that nearly always happens behind closed doors, despite the emergence of formal stipulated rules to achieving tenure. Chairs (positions) are increasingly rare, due to attributed shifts in funding discussed above.

In addition to these professional systems, there are also policies related to steering knowledge outputs through other means. These are often more specific, detailed lists of goals, or evaluation criteria. In many departmental units publication lists existed. The explicitness of these lists varied from formal to informal lists, though they were discussed and output goals were implied. These lists are used as targets for guiding and directing research.

In addition, in some cases departmental units also provided incentives for research achievements. These largely existed in being awarded more research time and less teaching and were given upon grant acquisition, or publication of acknowledged high quality. Since no explicit policies existed on these terms, incentives were largely informally organized by researchers. In a few cases, departments have individual teaching departments, meaning that individuals are hired with the sole purpose of teaching (See the Universiteit van Amsterdam Lecture Group as an example). In this case it is might also be a ”de-incentive”, for demotion with a larger education load is often given to those with low academic production.

The Dutch Computer Science departmental units have a number of policy mechanisms that attempt to steer individual researchers' behavior. In this dissertation research I focus on the rules, in the form of policies or targets, that attempt to steer publications outcomes. In exploring the effect of these policies as measurements of the rules, I am able to question how these rules provide a contingency for the facilitation of specific networking behavior by individual researchers. I present these here, not as propositions–as they are specified in greater details as propositions and hypotheses given the research questions explored in the separate chapters–but as a description of how identified factors of departmental unit descriptions relate to the study of network dynamics. For example, the existence of a professional tenure system aimed at facilitating specific behavior and outcomes in the case of knowledge-intensive science work may very likely stimulate or hinder the emergence of specific networking behaviors for the achievement of these publications.

## 2.8 Conclusion

In this Chapter, I presented a qualitative description of the case of Dutch Computer Science researchers drawn from findings from interviews combined with supportive tertiary material outline a number of key issues in the field. As I mentioned in the methods section, these findings are subject to the sampling method used and, thus, I acknowledge that it is a general overview of the field of computer science in The Netherlands.

Computer science researchers in The Netherlands are specialists in a variety of sub-fields; they conduct their work in a number of distinct organizational settings, with diverse opportunities given their embeddedness at a time of uncertain funding changes. This situation is related to the increasingly perceived autonomy of researchers in the production of knowledge. Such a qualitative portrait provides a cornerstone for understanding insights from empirical studies of this dissertation research on the emergence of specific scientific collaboration networks among Dutch Computer Science researchers.

# 3 Context, Network, & Performance: Contingencies of Successful Collaboration Networks[i]

## 3.1 Abstract

In this paper, we develop a theory for how network constraint  the extent to which an actor's alters are connected to each other  in individual collaboration networks affects performance. Using a combination of longitudinal bibliometric publication data and survey data, we analyze the scientific collaborations of 193 Dutch Computer Science researchers. We show that in such a knowledge intensive environment, adopting policies that steer outputs enhances the success of scientific collaboration in high constraint networks. These results contribute to scholarly knowledge about the effect of context on the success of networks with different levels of constraint; where constraint in knowledge intensive environments is enhanced by rules of the organization.

## 3.2 Introduction

The array of relationships within social networks allows individuals to tap into important resources (Burt, 2005; Coleman, 1988). This is especially relevant in light of the fact that the positions of individuals within social networks, as well as network structure, have consequences for outcomes (Ahuja, 2000; Burt, 2005; Coleman, 1988; Cummings and Cross, 2003; Fleming et al., 2007; Granovetter, 1985; Krackhardt, 1999; McFadyen et al., 2009; Mehra et al., 2001; Obstfeld,

[i]This paper is presented as a working paper in preparation for future journal submission. This paper was prepared in cooperation with Dirk Deichmann, Peter Groenewegen, and Edwin Horlings. Thus it is written in the plural we form.

2005; Tortoriello and Krackhardt, 2010). Scholars have long debated the benefits of constraint in networks - the extent to which an actor's alters are connected to each other. However, optimal network structures remain debated (Burt, 2005; Coleman, 1988; Granovetter, 1985). In an attempt to shed light on this puzzle, some studies have suggested that specific network structures provide benefits depending on the context of the networks. (Battilana and Casciaro, 2012; Burt, 2000).

When taking into account network context, the efficiency of network structures has been questioned when considering a number of these context conditions. For example, networks that are low in constraint facilitate performance in completing non-complex, less knowledge intensive tasks (Hansen, 1999; Krackhardt, 1992). Inversely, high constraint in networks has been shown to be most effective for completing complex knowledge tasks (Cummings and Cross, 2003; Hansen, 1999; Reagans and McEvily, 2003). These dense networks facilitate performance through increased identification in a group (Borgatti et al., 2009), and trust (Borgatti and Cross, 2003). Furthermore, Burt (2000) suggests that under conditions of high uncertainty and in small groups, low constraint (i.e. structural holes) is most beneficial for performance. A highly constrained network structure has been found to improve performance when work is performed under low uncertainty in larger groups.

The proposition that specific conditions shape the effect of network structures on performance is not new, although the role of the organization where individuals pursue their work has been arguably overlooked in understanding the emergence of specific network structures. Recent work has suggested that in organizations undergoing change the specific contextual conditions under which networks are invoked influence the success of different structures (Battilana and Casciaro, 2012). When organizations undergo changes that divert largely from the status-quo, they are generally more successful in cohesive networks with higher constraint. Consequently, the role of organizations where individuals pursue their work has an effect on specific network structure employed. Formal organizations, such as universities, often impose constraints and create opportunities through policies that seek to guide behavior and outcomes. In an attempt to elucidate the role of these policies in influencing network structures and subsequent outcomes, we consider the following: *which organizational policies influence networks such that individual performance improves?*

We specifically examine networks through the lens of knowledge-intensive collaborations among science researchers. In science, publication success plays an important role in career defining events such as promotion for tenure and grant acquisition. These successes, typically measured through citations (Wouters, 1999) are a recognition of quality and influence in any scientific field. Increasingly, this knowledge is produced in teams (Wuchty et al., 2007) as observed through the prevalence of increasing numbers of co-authors on academic publications (Greene, 2007); thus publication success cannot be attributed to individual factors alone. A researcher's relationships in scientific networks garner access to different information and expertise. Researchers may have greater publication success given the structure of their individual scientific collaboration network. Given this perceived benefit of collaboration (Hudson, 1996), scientific organizations, academic institutions and policymakers are seeking ways to facilitate and encourage collaboration through team science (Falk-Krzesinski et al., 2010), as well as through formal research initiatives such as grants and institutes (Defazio et al., 2009). In this paper, we aim to evaluate this aim and investigate the effect of organizational policies for collaboration on individual outcomes.

With this in mind, we define this context through the identification of specific policies that work to potentially stipulate the conditions of outputs such as, for example, a contractual stipulation on achieving tenure. We suggest an interaction between individual network structure and formal organizational, contextual effects on performance. We first review current theories and identify the potential role of policies in the success of specific network structures. We propose a number of hypotheses in which contextual factors function as influential contingencies in specific network structures. To test these effects, we collected data via a survey and interviews and queried bibliometric data on Dutch Computer Science researchers between 2006 and 2010. We then analyze how network configuration of co-authorship networks leads to publication success, depending on departmental output policies where the researcher is embedded. Findings not only contribute to existing theories that specify conditions under which network structures are successful, but advance our practical understanding of the effect of conditions established by organizations through policies that enhance the success of specific network structures.

## 3.3 Networks and Context

Network scholars have long recognized the consequences of network structures on performance. These organizational structures are related to the access granted to valued resources (Brass, 1984; Ibarra and Andrews, 1993). The extent to which an actor's alters are connected to each other is referred to as constraint. The effects of constraint is the extent to which an actor's alters are connected to each other positively influences performance have been long debated in social network studies. Constraint is considered from two extremes – structural holes (i.e., no or low constraint) and closure (i.e., high constraint). A structural hole exists when an actor, the ego, is situated between other actors who are not linked. Such network structures expose the ego to non-redundant information which facilitates brokerage (Burt, 2004).

In turn, such brokerage may lead to increased control and access to information which influences performance (Fleming et al., 2007; Mehra et al., 2001; Obstfeld et al., 2014). However, there is also evidence that closure in an individual's network enhances success (Obstfeld, 2005). Closure describes a dense, cohesive network structure (Coleman, 1988) that efficiently mobilizes resources and people through multiple connections where all actors have the ability to share and exchange knowledge (Coleman, 1988; Granovetter, 1985; Tortoriello and Krackhardt, 2010). Thus, both network structures are seen to have advantages.

Despite these views, the efficiency of network structures has been questioned, especially in relation to a number of conditions. Recent work has suggested that individual and contextual conditions play a role in the performance of networks. Networks that are low in constraint facilitate the performance of non-complex, less knowledge-intensive tasks (Hansen, 1999; Krackhardt, 1992). Inversely, high constraint in networks has been shown to be most effective for completing complex knowledge tasks (Cummings and Cross, 2003; Reagans and McEvily, 2003). Burt (2000) suggests that under conditions of high uncertainty and in small groups, low constraint (i.e., structural holes) is most beneficial for performance, while a constrained network structure has been found to improve performance when work is performed under low uncertainty in larger groups. Similar findings were confirmed by Battilana and Casciaro (2012) suggesting that change agents' networks are generally more successful in cohesive networks with higher constraint in organizations undergo changes that divert largely from the status-quo. Networks with less redundant contact better performing in new markets or with new technologies of uncertainty than in a familiar terrain where redundant contacts proved more influential to success (Burt, 2000; Hansen et al., 2001; Podolny, 2001). Burt et al. (2000) found differences in the success of the networks of managers between American and French companies, suggesting that the culture as an external factor plays a role. Additionally, differences in tasks and tasks uncertainty play a role in the success of networks; with. These external factors are said to play a greater role on the networking decisions of individuals than an individual's characteristics (Burt et al., 1998).

Thus, the proposition that specific conditions shape the effect of network structures on performance is not new, although the role of formal organizational context has been arguably overlooked in understanding network mechanisms. Due to the dominant assumption that individuals organize their network and form ties to other people autonomously and relatively independent of any contextual influence (Burns and Stalker, 1961; Daft, 1982) the role of formal organizational context has been arguably overlooked in understanding network mechanisms. Thus, context can be taken into account by acknowledging for instance that formal organizational bodies provide rules, policies, and regulations that facilitate the success of different network structures.

We aim to further elucidate the role of the formal organization through investigating policies implemented by formal organizational bodies that outline a specific set of their employees' core tasks, many of which are often contractual. These formal conditions stipulate a process in which an employee is evaluated and where the process of promotion, firing and hiring is transparent. Such a policy does not guarantee a position but instead mandates due process and outlines how to obtain a specific status. For example in teaching and academic professions, a professional is on a tenure track, or the conditions for a lawyer to make partner. In addition to these professional systems, there are also day-to-day conditions stipulated by the formal organization that facilitate the employees' work. These are often more specific and detailed lists of goals, or evaluation mechanisms that attempt to steer specific behaviors or outcomes.

Given this, we postulate the following: Considering that knowledge is by nature codified and tacit (Latour and Woolgar, 1979), knowledge-exchange would therefore be greatly facilitated in a network structure of high constraint. The reduction of exchange barriers via multiple shared relationships in a network increases understanding within a cohesive group of individuals and thus yields greater success in terms of outcomes. Within knowledge-intensive activities, the space that individuals have to navigate policies potentially alters the way in which individuals might approach potential collaborators for future projects. The professional tenure system aims to facilitate specific behaviors and outcomes; in the case of knowledge-intensive work described above, this would aid in stimulating cohesive, closed network structures, reduce barriers and thus increase success. In other words, the effect of constraint on performance is enhanced by policies that attempt to stimulate specific behavior in researchers as they undertake their work. An evaluation policy such as a target list creates heightened awareness of a specific outcome, further stimulating a constrained cohesive network structure that reduces barriers for more effective knowledge exchange. The success of a constrained network is enhanced when both policies exist within an organization. Thus we propose:

**Hypothesis 1:** A highly constrained network will be strongly associated with publication success in an organization with a professional tenure system, as compared to a low constrained network.

**Hypothesis 2:** A highly constrained network will be strongly associated with publication success in a department with a target list, as compared to a low constrained network.

**Hypothesis 3:** A highly constrained network will be strongly associated with publication success where both a professional system and a target list exist, as compared to a low constrained network.

## 3.4 Method

### 3.4.1 Setting

We investigate our hypotheses in the setting of academic science and investigate the success of scientific collaboration via researcher co-authorships. Researcher relationships in scientific networks help garner access to different information and expertise. Scientific collaboration provides a unique setting for the exploration of these network structures, as collaborative relationships are largely established through individual researchers' affiliation with specific departments that have varying combinations of policies toward the guidance of behavior.

Co-authorship, a standard proxy for scientific collaboration (Melin, 2000), provides a valid and reliable source for measuring both networks and outcomes though bibliometric databases. Additionally, there is a large body of scientific work exploring scientific collaboration. The manner in which researchers employ different network structures in the development of published work has most often been studied at the individual level, where researchers' collaboration tendencies and performance are explained from the point of view of their formal position, gender, and experience (Bozeman and Corley, 2004; Lin and Bozeman, 2006). Recent studies, however, suggest that institutional factors (Ponomariov, 2007; Ponomariov and Boardman, 2010; Shrum et al., 2007) and the demands of funding schemes (Defazio et al., 2009) also influence collaboration.

Science provides a number of clear organizational entities for the consideration of potential rule makers in the measurement of contextual factors. These include practices within the discipline (Whitley, 2000), *supra*national science policies and regulations, institutional policies, and departmental polices. These bodies implement policies and regulations in an attempt to guide behavior. Academic institutions, and departments in particular, provide a residence for researchers to conduct their work, and they attempt to steer knowledge processes through the facilitation and constraint of different ideal behaviors through policies and incentives (Fairweather, 2002; Fairweather and Beach, 2002). One way institutions achieve this is via evaluation policies , two policies in particular: a formal professional tenure system and a publication list. First, with a tenure system, a department stresses that a certain amount of papers need to be published in a given time frame for a scientist to receive or continue to have tenure. This is a contractual obligation that has consequences; thus it is in the researcher's interest to most effectively produce high quality papers. Those scientists that have a constrained social network profit from this policy as they have reduce barriers to production. Ultimately, they will therefore publish papers that receive more citations and that are therefore argued to be more successful.

Organizational dynamics vary across fields (Garg and Padhi, 2001); thus when selecting a population to examine, one field was selected, as well as one national context, in order to limit variance in the external factors affecting scientists. In this study we investigate Dutch Computer Science. The Dutch context was selected for it is a typical European academic environment with funding on the national level for research-stimulating cooperation; it provides a diversity of cases at which to examine different institutional processes in a relatively small geographical space. The field of computer science was chosen for two reasons: the high propensity of collaboration through co-authorship and the validity and reliability of online sources for documenting publications.

Computer Science is a field based on both information and computation studies, coming together in the use of computers as systems and or tools for solving research problems. The discipline of computer science is a mature, intellectually unified field with a number of mature sub-fields existing as self-sustaining practices. Computer Science subjects range from bioinformatics, artificial intelligence/cognitive science, cybernetics, quantum computing and business applications. Consequently the field not only works on internal questions but has a tendency to work with other fields.

Within the Netherlands, the discipline of computer science is a field of high research quality, with significant, though decreasing, funding for discipline-related work (Nationale Informaticakamer, 2010). To increase validity, we follow the selection made by the Nationale Informaticakamer's five-year review of Dutch academic research institutions as to contextualize with a formal public review on organizational and research processes; the following academic research universities have been selected: Universiteit van Amsterdam, Universiteit Utrecht, Technische Universiteit Delft, Vrije Universiteit Amsterdam, Technische Universiteit Eindhoven, Universiteit van Twente, Universiteit Leiden, Radboud Nijmegen Universiteit, and Rijksuniversiteit Groningen.

### 3.4.2 Sample

In this study, we look at individual scientific collaboration behavior. We start from a list of known active scientists. This list is compiled from two sources:

(1) a list of permanent staffed researchers in 2010 provided by NARCIS (NARCIS, 2014), and

(2) a manually collected list of all staff working at the 9 above mentioned institutions as listed on the affiliated institutional websites.

From this list publication data was queried from *Digital Bibliography & Library Project Database* (DBLP) (Database Systems and Logic Programming, 2013). DBLP is one of the most comprehensive bibliographic indices for the field of Computer Science and thus allows the identification of a valid and reliable set of co-authorship publication data for Dutch Computer Science researchers. This query resulted in a list of both Dutch researchers and their co-authors. In order to confirm affiliations of the researchers and increase the validity of the identified list, the institution was identified through a query of two additional databases. The automatic collection of historical data on institutional affiliation is not currently stored in one database. A query using Microsoft Academic Search – a database which includes the DBLP data set – was used to identify institutions (Microsoft Academic Search, 2014). To locate additional missing data, another database, ArnetMiner.org (ArnetMiner, 2012) – a search and mining service of Computer Science researchers which includes semantic data on computer scientists – was used. In order to disambiguate institutional names this list was further queried in geocoding Web service Yahoo! PlaceFinder(Yahoo! PlaceFinder, 2012). We identified 1516 Dutch Computer Science researchers.

An online survey was used to collect data on the individual researchers. Dutch Computer Science researchers were asked to reflect on their position, background and attitudes. The online survey tool – Qualtrics – was implemented in this research. Qualtrics provides a user-friendly survey design which makes it possible to directly transfer answers into a database. The survey, Computer Science in The Netherlands was distributed via email to the source list of Dutch Computer Scientists in week 17 of 2013. Two reminders were sent two weeks after the first email, and again four weeks after. All 1516 scientists were emailed and asked if they were affiliated with the Computer Science department of interest in 2006 – 2012. By taking a snapshot of researchers' collaborations during this period, we are able to identify a specific period of time at which explore the effects of the policies. This method allowed us to filter out researchers who were perhaps listed on the website but formally worked elsewhere and thus were not directly or daily influenced by the potential organizational context of interest. This effort resulted in 214 responses – a 14% response rate – of which 193 responses were complete.

### 3.4.3 Model

We test our model using quantitative and qualitative data from scientific collaborations among researchers considering context, network and performance. We measure context through the identification of policies of academic departments. We consider ego networks of scientific collaborations through co-authorship to assert network structures. We look at the performance success of these specific collaborations by considering citation impact and a set of control variables related to individual researchers.

### 3.4.4 Measures

**Publication success** Citations are defined mentions from one scientific publication within another which recognizes other knowledge (Weinstock, 1971). The number of citations indicates the quality or value of a piece of knowledge in related field. It is consequently an important part of the scientific evaluation process (Radicchi et al., 2008). Citation records per Dutch Computer Science researcher within the set was acquired from the bibliographic database Microsoft Academic Search (Microsoft Academic Search, 2014). This is a valid database to extract citation records as the entire DBLP publication dataset, where we queried the publication data as embedded in this database in embedded in MAS, thus insuring no missing data. This database was queried in summer 2013, providing raw citation scores for all the related publications from 2006 – 2012, allowing a lag period for citations.

**Constraint** Co-authorship is viewed as a valid and reliable measure of collaboration and assumed interaction for knowledge production, as credit is given to those involved in research (Hicks and Katz, 1996; Melin and Persson, 1996). We use publication data to define the scientific collaboration network of the researcher. Networks are inferred through co-authorship. The field of computer science has a number of internally managed publication databases. These databases allow us to make a valid selection of publications from our sample population, compared to the use of Web of Science which has acknowledged biases for the field (Bar-Ilan, 2010b). The researchers names were queried in DBLP for all historical publication data. From these data, we made a selection of a period of four years (January 2006 – December 2012). This time period corresponds to the recall-ability of answers about collaboration in the survey and also considers a publication lag period for the emergence of collaboration efforts within the organizational policies.

Ego networks were investigated for each individual author consisting of authors as nodes and ties as shared publications over the six year period. We measured the constraint in the researchers' collaboration network using Burt (1992). Constraint measures three dimensions of the network: network size (larger networks are less constraining), density (networks of more strongly interconnected contacts are more constraining), and hierarchy (networks in which all contacts are exclusively tied to a single dominant contact are more constraining) (Burt, 2004). This allowed us to capture a continuous measure of constraint on the network structures thereby providing insight into the balance between structural holes and cohesion at which to test our hypotheses. The analysis of all 193 ego networks was completed in UCINET (Borgatti et al., 2002).

**Context: Professional system and target list** To specify the context we identify policies of the Dutch Computer Science departments. Semi-structured interviews with the heads of the nine academic departments were completed by the first author. In two cases the department heads were not available or not willing to be interviewed; they thus recommended someone in their place. In one case this was the former department head and a head of education within the management team. Policies identified in these interviews included professional tenure systems, formal cooperation with industry, grant-acquisition strategies, lists of publications, and incentives via less teaching time for grant acquisitions. In this study we focus on policies that attempt to steer the behavior of researchers in regards to their publications. Thus we identify the existence of a professional tenure system and the existence of a publication list. It is not always the case that publication lists are embedded in formal evaluation systems as often these tenure systems are regulated by the university itself and thus just demand high quality that can be validated. Thus we identify the existence of tenure systems at any time between 2006 and 2010, as indicted from the interviewees.

In identifying the existence of a publication list we asked survey respondents to reflect on a set of conditions within the department. This allowed us to determine whether departments kept lists of publications; such lists then allowed us to determine whether sub-departments or groups had specific goals.

**Controls** Science studies scholars have long recognized how the effect of individual characteristics, positions, geographical proximity, and experience influence collaboration (Boardman and Ponomariov, 2007; Bozeman and Corley, 2004; Bozeman and Gaughan, 2007; Corley et al., 2006; deB Beaver and Rosen, 1979; Liben-Nowell and Kleinberg, 2007; Newman, 2004, 2001a,b; Ponomariov, 2007; Ponomariov and Boardman, 2010; Rodriguez and Pepe, 2008; Shrum et al., 2007; Stokols et al., 2008a). The propensity to collaborate is also related to scientists' attributes, e.g. gender, tenure, and field experience (Bozeman and Corley, 2004; Bozeman and Gaughan, 2011; Cole and Zuckerman, 1984; deB Beaver and Rosen, 1979; Melin, 2000; Stokols et al., 2008a). Keeping this in mind, we considered a number of researcher characteristics as controls, such as gender, nationality, position, subfield, affiliation with research institute, experience outside academia, number of solo publications, number of publications before 2006, number of publications after 2006, number of co-authors before 2006, number of co-authors after 2006, and citations before 2006 of the individual ego networks. This control data was collected via a survey and was computed from the bibliometric data on the individual scientists in order to control for alternative explanations in the understanding of the enhanced effect of organizational context on the success of network structures, as hypothesized above. We log-transformed all count variables.

### 3.4.5 Analysis

We used multilevel modeling to take into account the nested data structure: employees on level 1 and department groups on level 2. We used Stata 13.0 to run mixed-effects linear regressions on our log-transformed variable citations (Rabe-Hesketh and Skrondal, 2008). The constraint variable was mean-centered before interactions with this variable were entered into the different models.

Table 3.1: Models

| Variable | Publication Success (citation score) Model1 | Model2 | Model3 | Model4 | Model5 |
|---|---|---|---|---|---|
| Constant | -0.48 | 0.16 | -0.01 | 0.06 | 0.50 |
| | (0.66) | (0.66) | (0.68) | (0.70) | (0.70) |
| Position | 0.00 | -0.03 | -0.03 | -0.02 | -0.01 |
| | (0.07) | (0.07) | (0.07) | (0.07) | (0.07) |
| Gender | 0.06 | 0.13 | 0.09 | 0.12 | 0.12 |
| | (0.22) | (0.21) | (0.21) | (0.21) | (0.21) |
| Dutch nationality | -0.14 | -0.14 | -0.14 | -0.15 | -0.18 |
| | (0.17) | (0.16) | (0.16) | (0.16) | (0.16) |
| Subfield applied mathematics | 0.06 | 0.56 | 0.64 | 0.62 | 0.19 |
| | (0.71) | (0.70) | (0.71) | (0.71) | (0.71) |
| Subfield artificial intelligence | 0.17 | 0.55 | 0.55 | 0.50 | 0.14 |
| | (0.61) | (0.60) | (0.60) | (0.60) | (0.61) |
| Subfield bioinformatics | -0.37 | 0.07 | 0.12 | -0.10 | -0.44 |
| | (0.75) | (0.74) | (0.74) | (0.75) | (0.74) |
| Subfield traditional computer studies | 0.12 | 0.46 | 0.49 | 0.42 | -0.00 |
| | (0.62) | (0.61) | (0.60) | (0.61) | (0.62) |
| Subfield databases management | 0.13 | 0.60 | 0.61 | 0.65 | 0.31 |
| | (0.69) | (0.68) | (0.68) | (0.68) | (0.68) |
| Subfield image sound | 0.20 | 0.51 | 0.49 | 0.42 | 0.04 |
| | (0.70) | (0.68) | (0.68) | (0.69) | (0.69) |
| Subfield web development design | -0.31 | 0.03 | 0.00 | -0.08 | -0.49 |
| | (0.75) | (0.73) | (0.73) | (0.73) | (0.74) |
| Subfield networks | 0.24 | 0.61 | 0.61 | 0.63 | 0.24 |
| | (0.64) | (0.62) | (0.62) | (0.63) | (0.64) |
| Subfield operating systems | 0.26 | 0.65 | 0.69 | 0.68 | 0.28 |
| | (0.77) | (0.75) | (0.75) | (0.75) | (0.75) |
| Subfield simulations | 0.15 | 0.51 | 0.57 | 0.48 | 0.12 |
| | (0.62) | (0.61) | (0.61) | (0.61) | (0.62) |
| Subfield software systems | 0.44 | 0.80 | 0.80 | 0.76 | 0.36 |
| | (0.62) | (0.61) | (0.61) | (0.61) | (0.62) |
| Affiliation with research institute | -0.07 | -0.02 | -0.02 | 0.03 | 0.04 |
| | (0.16) | (0.15) | (0.15) | (0.15) | (0.15) |
| Experience outside academia | 0.24 | 0.19 | 0.18 | 0.17 | 0.19 |
| | (0.15) | (0.14) | (0.14) | (0.14) | (0.14) |
| Number of solo publications (ln) | -0.04 | -0.02 | -0.04 | 0.03 | 0.03 |
| | (0.18) | (0.18) | (0.18) | (0.18) | (0.18) |
| Number of publications before 2006 (ln) | -0.14 | -0.12 | -0.11 | -0.19 | -0.19 |
| | (0.19) | (0.18) | (0.18) | (0.18) | (0.18) |
| Number of publications after 2006 (ln) | 0.54*** | 0.67*** | 0.69*** | 0.70*** | 0.75*** |
| | (0.14) | (0.14) | (0.14) | (0.14) | (0.14) |
| Number of co-authors before 2006 (ln) | -0.18 | -0.15 | -0.16 | -0.11 | -0.12 |
| | (0.16) | (0.16) | (0.16) | (0.16) | (0.16) |
| Number of co-authors after 2006 (ln) | 0.34* | -0.15 | -0.12 | -0.15 | -0.21 |
| | (0.14) | (0.20) | (0.20) | (0.21) | (0.21) |
| Citations before 2006 | 0.79*** | 0.76*** | 0.76*** | 0.78*** | 0.76*** |
| | (0.05) | (0.05) | (0.05) | (0.05) | (0.05) |
| Constraint | | -1.83*** | -1.77*** | -1.87** | -1.84** |
| | | (0.52) | (0.52) | (0.67) | (0.66) |
| Target list | | | 0.16 | 0.17 | 0.17 |
| | | | (0.17) | (0.17) | (0.17) |
| Professional tenure system | | | 0.19 | -0.03 | 0.02 |
| | | | (0.25) | (0.29) | (0.29) |
| Constraint x Target list | | | | -0.53 | -0.83 |
| | | | | (0.66) | (0.65) |
| Constraint x Professional tenure system | | | | 1.92* | 1.35 |
| | | | | (0.92) | (0.93) |
| Target list x Professional tenure system | | | | 0.30 | 0.74 |
| | | | | (0.57) | (0.58) |
| Constraint x Target list x Professional tenure system | | | | | 10.23** |
| | | | | | (3.92) |
| Variance of constant | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Variance of residual | 0.94 | 0.88 | 0.88 | 0.85 | 0.82 |
| Log likelihood | -267.89 | -261.81 | -261.03 | -257.90 | -254.56 |
| Observations | 193 | 193 | 193 | 193 | 193 |
| Number of groups | 9 | 9 | 9 | 9 | 9 |

Standard errors in parentheses
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, $+p < 0.10$

## 3.5 Results

The descriptive results and correlations of all the variables considered in the models can be seen in the Appendix, Table A.2. Correlation coefficients are marked statistically significant ($p < .05$), there is a negative correlation between constraint and performance success ($r = -.55$, $p < .05$). Table 3.1 displays the the results of the mixed-effects linear regressions. Model 1 includes control variables likely to influence researchers' performance. Results show that the number of publications after 2006 ($\beta = .54$, $p < .01$), number of co-authors after 2006 ($\beta = .34, p < .05$), and citation performance before 2006 ($\beta = .79, p < .001$)have a positive and significant effect on publication success. Model 2 introduces the constraint of the individual scientific collaboration networks, which measures the degree of structural closure in the ego network. This effect is negative and statistically significant ($\beta = -1.82$, $p < .001$). This means there is a positive relationship between the likelihood of structural holes and publication success.

In Model 3 the variables professional tenure system and target list are added. Both are not significant. Model 4 introduces the multiplicative term for constraint and the output policies professional tenure system and a target list. The interaction effect of constraint and professional tenure system is positive and significant ($\beta = 1.92$, $p < 0.05$). in which a highly constrained network will be strongly associated with performance in an organization with a professional tenure system. We cannot confirm Hypothesis 2 as the effect of a publication list in the department was not a significant moderator of the relationship between constraint and publication success.

Model 5 introduces the three-way interaction which is positive and significant at the ($\beta = 10.23$, $p < 0.01$). This suggests an effect of moderation on the part of a professional tenure system and a target list in enhancing the success of constrained scientific collaboration networks of researchers. This confirms Hypothesis 3 that a highly constrained network will be strongly associated with performance under conditions of both a professional system and a target list. These findings confirm that the effect of network closure on publication success is enhanced in departments with policies stipulating rules for producing publications.

### 3.5.1 Discussion

Findings indicate that the level of constraint in researchers' social networks is contingent on the policies of their respective departments. If the department has a publication list and a tenure system, researchers have more success if embedded in a highly constrained network. However, if neither publication list nor tenure system exist, researchers perform better in a network with low constraint. If the department has a publication list, researchers have enhanced success with networks that are constrained; but if there is no such list, they have higher success with a network that has structural holes (low constraint). If the department

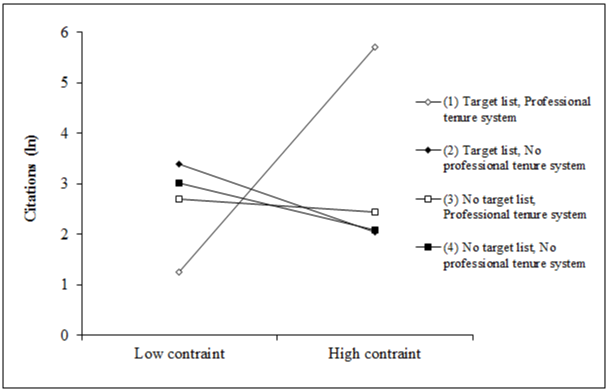



Figure 3.1: Interaction

has a publication list but no tenure system exists, performance is also higher in networks with low constraint; when no formal list exists but a tenure system does exist, the same holds true, and performance is higher in networks with low constraint.

More generally we purport that the conditions under which researchers develop and maintain their scientific collaboration networks influences the success of a network. These output policies serve as a sort of herding mechanism, working to facilitate specific types of networks of researchers. These are distinct where success in a network with low constraint allows the researcher to bridge or manage a number of clusters of collaborators, compared to success in a network that is more cohesive (high constraint), where communication barriers are reduced. When a department has an evaluation list and a tenure system, it provides certainty about evaluation criteria. We suggest that this leads to distinct networking strategies where those in departments with specifications researchers can focus on their local position (i.e. within the department) and investing in cohesive teams, whereas without a list or tenure track system, performance criteria are uncertain and researchers must position themselves more globally. Strategic position in one context requires a different network than in the other. Researchers are facilitated and/or constrained by these specific policies about output and thus strategize accordingly. While the lack of policies yields similar performance by researchers as that achieved when both lists and tenure are present, very different strategies for networking result. A mix of policies yields, overall, lower performance than when policies are explicit, which perhaps identifies a type of uncertainty within the context and leads to more experimentation with potential co-authors.

This study provides clear support for a causal theory that stipulates the interaction of organizational policies on the success of network structures. Findings suggest that, depending on departmental policies, constraint plays different roles in the success of individual researchers' scientific collaboration networks. The policies of a professional tenure system and a target list within a department enhance the success of cohesive ego networks scientific collaboration.

Thus, organizational context is an important enabling or constraining condition in the success of different network structures. The results from the current study suggest that knowledge intensive organizations (e.g., universities) have the means via policies to affect the success of collaboration networks among scientists. These results are particularly important for individual scientists and universities. First, individual scientists can try to build their collaboration networks in accordance with the departmental policies. For example, if a researcher works in a department which has both a publication list and a tenure system, a highly constrained collaboration network would be optimal. This means that the researchers' collaborators are also relatively well connected among each other. If, on the other hand, the department has neither publication list nor tenure system, our findings suggest that the researcher might better build a network saturated with structural holes, e.g., low constraint. This means that the researchers' collaborators are not particularly well connected among each other. Second, universities can build on the insights of this research by streamlining their policies for collaboration and evaluation. For example, if a department has a publication list and a tenure system, it might pay off to encourage ties between collaborators by building opportunities for these collaborators to meet more often. Conversely, if a department has neither publication list nor tenure system, it might be best for performance of individual researchers to promote and encourage contacts with new collaborators, through e.g. facilitating conference visits. Concluding, both universities and researcher will profit from more attention to the contingencies of the social networks that they are embedded.

These findings contribute to the ongoing debate on constraint between the efficiency of structural holes and cohesive networks (Burt, 2005; Coleman, 1988; Granovetter, 1985), through specifying the conditions under which specific networks are successful. This research confirms previous studies that argued that high constraint in networks is most effective for completing complex knowledge tasks (Cummings and Cross, 2003; Hansen, 1999; Reagans and McEvily, 2003). However, our findings question Burt's (2000) proposition that constraint is an advantage in larger groups with low uncertainty. Specifically, within knowledge-intensive activities such as scientific collaboration, policies that attempt to avoid uncertainty alter the way in which researchers realize publications via potential collaborators. Future research should explore the conditions under which this holds in both less knowledge intensive contexts and other non-research networks to further confirm these findings. Future work should also explore additional performance measures that are non-peer-review related, as to increase possible impartiality of performance.

# Considering context: recasting the SIENA model to consider contextual factors as determinants of social network structures[j]

# 4

## 4.1 Preface

The previous chapter investigated context in considering a network consequence question, thereby advancing knowledge on the effect of context in understanding the success of network structures. This chapter, and the one following aims to build on information about how context serves as a determinant for the emergence of network structures, providing insight into the antecedents of these scientific collaboration networks.

## 4.2 Abstract

Advances in knowledge regarding social network dynamics have resulted in increasingly complex techniques for the investigation of network dynamics. Naturally, each of these methods has its limitations, stemming from underlying theoretical assumptions. Nonetheless, a number of gaps exist in current knowledge which cannot be accounted for within current network methodological specifications, specifically the role of exogenous contextual factors – the setting or environment. Through a review of current methods we identify a blind spot in the exploration of the effect of context as a determinant for the generative

[j]This paper is presented as a working paper in preparation for future journal submission to Organizational Research Methods and was prepared in cooperation with Dieuwke Ydel, Rena Bakhshi and Peter Groenewegen. Thus it is written using the plural we form.

mechanisms that explain network structures. We prose a recasting of the use of SIENA, a simulation-based network model which qualitatively compares models to understand the effect of different contextual factors on the mechanisms that explain observed network structures. We test this methodology on individual scientific collaboration patterns of Dutch Computer Science researchers within several academic departments. Our findings show that exogenous contextual factors influence the strength of the observed generative mechanisms utilized within researchers' networks. Our results suggest that these factors play a stronger role in network dynamics than currently theorized and that such a method provides distinct insight with which to advance on both empirical and theoretical levels for the investigation of social network dynamics.

## 4.3 Introduction

A *network* refers to any organizational form with a distinct structural form (Powell, 1990). In particular, a social network is one that represents relations between actors (Borgatti et al., 2009). These relations or set of ties define a network type, such as a friendship tie in a friendship network or an advice tie in an advice network. The structure of the social network changes through the creation, maintenance and dissolving of relations between actors. Advances in knowledge on social network dynamics have increased rapidly in the last decade with increasing sophistication of techniques and emerging theories of these network processes.

Theories regarding the emergence of social network structures suggest that multiple generative mechanisms govern network formation (Monge and Contractor, 2003). These mechanisms are classified into different roots (Albert and Barabási, 2002; Blau, 1977; Ibarra, 1992; Katz, 1953; Liben-Nowell and Kleinberg, 2007; McPherson et al., 2001; Monge and Eisenberg, 1987). The roles these roots play in generating networks remain debated, with theories suggesting different *modi operandi* (modes of operation). Context is overlooked as an explanatory variable in network dynamics. This is due to the ontological lens of the past 30 years of social science from which this knowledge has emerged as a topic. This perspective seeks to understand the individual and the individual's interactions and behavior within the network. Thus network processes are said to be driven (largely) by the individual. This is mirrored in the methodological tools used to investigate dynamics, where external factors are considered boundary conditions. This consideration poses a theoretical limitation and has consequences for the development of methodological tools. This is so because it limits the empirical testing of exogenous mechanisms as a possible explanatory variable for the mechanisms that generate network structures. Thus, we question: *How can we consider exogenous contextual factors as a determinant to the emergence of generative mechanisms that explain network structures?*

**Contribution and the outline** We begin in Section 4.4 outlining the theory on social network dynamics. We present and explain a gap in specifying how exogenous contextual factors exist as contingencies to the emergence of roots. We present the three most commonly used methods in social networks dynamics, further highlighting limitations in understanding exogenous contextual effects. In Section 4.5, we propose a method for considering contextual effects as a determinant for the emergence of mechanisms that can be realized given current methods. We consider the network of Dutch Computer Science researchers within nine academic departments for the exploration of this methodology. Our results, in Section 4.6, show that context influences the emergence of observed generative mechanisms of researchers' networks within academic departments, suggesting that these factors play a contingent role in network dynamics. In Section 4.7, we conclude with a number of considerations for expanding current methodologies used in social science network studies.

## 4.4 Theory

Theories on the emergence of social network structures suggest that multiple generative mechanisms govern network formation (Monge and Contractor, 2003). These network mechanisms have three roots:

(1) endogenous network factors,

(2) individual factors, and

(3) contextual factors (Lusher et al., 2012).

Endogenous factors are attributed to studies emerging from physics which look at mechanisms with preferential attachment (Albert and Barabási, 2002; Katz, 1953; Liben-Nowell and Kleinberg, 2007) and small-world networks (Watts, 1999). Studies in the social sciences have attributed a number of mechanisms rooted in an actor's characteristics, as well as characteristics of the dyad itself that the actor is a part of, such as principles of homophily or reciprocity (Blau, 1977; Ibarra, 1992; McPherson et al., 2001). Contextual factors have been attributed to a number of mechanisms related to the common proximity between two actors which influences interaction and the tendency to building potential relationships (Borgatti and Cross, 2003; Monge and Eisenberg, 1987; Owen-Smith et al., 2002).

The roles these roots play in generating networks remain debated, with theories suggesting different *modi operandi* or modes of operation. One set of theories proposes contingencies on the *modus operandi*, suggesting that the root invoked is dependent on the type of relationship under study (Shumate and Contractor, 2013). At the same time an evolutionary perspective is advanced which purports that the role of the different roots is dependent upon the stage of development of the network (Poole and Contractor, 2011).A second set of theories proposes that specific roots have precedence in manifesting network structures: one suggests a strong role for internal structuring (networks that are reinforced by both endogenous network factors and individuals' characteristics) (Ahuja et al., 2012; Whitbred et al., 2011), and a second related to the dominance of endogenous mechanisms (Liben-Nowell and Kleinberg, 2007) emphasizes other factors that play a less significant role in the emergence of structures.

Table 4.1: Roots in different theories

| Theories / Roots | Network evolution Physics | Internal network processes | Type of relation | Evolution |
|---|---|---|---|---|
| Endogenous network factors | ✓ | | ✓ | ✓ |
| Individual characteristics | | ✓ | ✓ | ✓ |
| Exogenous contextual factors | | | ✓ | ✓ |

In Table 4.1, as it is displayed, we see how current theories overlook exogenous factors. This is striking given that context, identified as national or regional culture (Monge and Eisenberg, 1987; Owen-Smith et al., 2002), the organizational structure, working conditions or demands (Balkundi et al., 2007; Danowski and Edison-Swift, 1985; Shah, 2000; Tichy and Fombrun, 1979), is related to network processes and structures. Additionally, recent research on consequences has suggested that the success of network structures is related to a number of contextual factors (Battilana and Casciaro, 2012; Burt, 2000; Burt et al., 2000; Hansen et al., 2001; Podolny, 2001). Thus, there is increasing evidence from network studies that context mediates the success – and thus emergence – of network structures.

Despite these findings, the role of context remains understudied. This is due to the dominance of the current ontological lens of social science, from which this knowledge emerged as a topic. This dominant perspective seeks to understand the individual's interaction and behavior within the network as driven by the individual. It thus overlooks external factors (McEvily et al., 2014). Additionally, this perspective is reflected in the methodological tools used to investigate dynamics, where external factors are considered as fixed or controlled boundary conditions related to the consideration of populations from which phenomena is generalized (e.g. networks of a specific industry (Uzzi, 1997), classroom (Snijders et al., 2010), or disciplinary field (Newman, 2004)). Thus, the role of exogenous contextual factors as factor in the *modus operandi* of networks is a blind spot in the understanding of the emergence of social network structures. In addressing how to consider context as a determinant to the emergence of structures we yield to considering current dynamic social network methods.

### 4.4.1 Current methods

A number of methods are used to investigate network dynamics; that is the emergence, maintenance or dissolution of ties between actors. In this study we focus our exploration on the use of current methodological tools and how context influences the emergence of other generative roots. In other words, can the differences in emerged network structures be attributed to the moderating role of context in constraining or facilitating generative network mechanisms? I outline here commonly used models in the study of dynamics and suggest ways these can be used to investigate context as a determinant of the generative mechanisms that explain the emergence of the observed network structure.

The majority of studies in social sciences are qualitative and compare network relations with measurements from two or more sets of actors (cf. (Faust and Skvoretz, 2002) for a review). In these cases multiple time slices of a network are taken, and characteristics of the network change are described in order to identify notable differences in structures, patterns of relations, and makeup. In this research, I focus solely on computational methods used in the social sciences to identify patterns in networks. Computational network methods are distinct from traditional statistical methods as they consider interdependencies between individuals. The goal of such methods is to explain, through statistical models, the likelihood of relations between actors. Three main techniques are used in social network dynamics studies for this purpose[k]:

(1) Quadratic Assignment Procedure (QAPs) (Martin, 1999),

(2) multiple network modeling using the p* model (Anderson et al., 1999), and

(3) the SIENA model (Snijders et al., 2010).

The Quadratic Assignment Procedure (QAPs) was devised for tie prediction based on covariates influence on structural outcomes. Structural elements are measured over time and evaluated through a regression (considering network interdependencies). This method is most often used in questions of multiplexity, which seek to explain dynamics through an understanding of reinforced relationships. Consequently, for QAP, any type of multi-rooted/level data, as considered in this study, is a limitation because it means mixing up different distributions. In these studies, single bounded networks are evaluated to identify factors that explain the emergence of network structures.

A second model is the p* model, which simulates the likelihood of tie formation in multiple network structures (Anderson et al., 1999). The p* model is a tie-oriented model that seeks to predict the likelihood of tie emergence based on

[k]We also acknowledge the rise of agent simulations in studying network dynamics. Many of these agent simulations implement similar, if not the exact, mathematical processes for predicting network dynamics (e.g. Markov Chain models). Reviewing these models is beyond the scope of the current paper, and thus we focus here on the methods that are most often used in social network dynamics by social scientists in particular.

the fact that new ties emerge as a result of the existence of previous ties. It uses Markov random graph models to identify structural patterns which signify that a specific mechanism is taking place. This method provides a tool for identifying mechanisms within networks, given a single cross sectional data set. The p* model does not consider the order of networks of relations, rather the same value is given to all networks in order to identify network mechanisms. An extension of the p* model which allows a comparison of different network tendencies could accurately compare differences in multiple networks, where coefficients are normalized to delineate the differences among significant mechanisms (Anderson et al., 1999). A benefit of this model is the possibility of considering large social networks (greater than 1000 nodes), as well as the fact that mechanisms can be identified from cross-section data of just one capture of a network.

The third most commonly used method in social network analysis dynamics studies is the use of Simulation Investigation for Empirical Network Analysis (SIENA), which provides a model for the investigation of dynamics in actor oriented networks (Snijders et al., 2010). Actor-oriented models give priority to individual factors in explaining network emergence. It is a model suited for closed networks that simulates the likelihood of network change based on:

(1) data on relations of individuals within the closed network from two or more time periods, and

(2) a set of variables, parameters and/or covariates that theoretically influence a possible change in the network.

It simulates all possible networks from the first given period, assuming that every individual has the potential to make a tie with the other individuals in the network, and then calculates the likelihood that the given parameters influence the change observed in the second (and or consecutive) networks, using Markov chain modeling (Ripley et al., 2011). The SIENA model is suited for relatively small ($< 1000$ nodes) closed networks, given the large state space needed to calculate the Markov Chain model.

In all three models, exogenous factors are most often considered boundary conditions, by which the network investigated is the context. In some cases the exogenous factor is considered a dyadic parameter and/or variable (e.g. two nodes share an affiliation). However, SIENA is the most suitable model for addressing the role of different mechanisms, and thus roots, in understanding network dynamics. The assumptions of actor-driven dynamics of Markov Chain models implemented in SIENA are best suited for investigating multi-rooted dynamics, therefore. These models are dynamic as they consider previous states in simulating the likelihood of future states. SIENA models are suited for relatively small (approximately 1,000 nodes) closed networks (where all individuals know one another and their behaviors), given the large state space needed to calculate the Markov chain model. The statistical comparison of Markov chain models through normalization of coefficients, as done in the p* models, is not applicable to SIENA due to the fact that in SIENA statistics are inferred from the distinct starting points of these models (Lusher et al., 2012). A number of mathematical solutions exist for comparing Markov chains (Müller and Stoyan, 2002), although few have been applied in social science studies due to the sensitivities of these simulations on starting data. It should be noted that multilevel models exist within the SIENA framework which yield typical networking tendencies given a sample of a population considering sequentially within and between-network analysis. However, the assumption remains that individual networks limited by an implicit boundary  a classroom, a school, an organization – are aggregated to say something about the population (Snijders and Baerveldt, 2003). Consequently, a methodological adaptation of SIENA would be necessary to consider exogenous contextual factors as moderators to the emergence of different generative mechanisms that explain the network structures. Thus, we propose a recasting of the current model, which allows us to profit from the well-tested, valid method of SIENA, explained below.

### 4.4.2 Recasting

We propose a recasting of the typical use of the SIENA model to assert a possible contingency of exogenous contextual factors on network dynamics, where models are qualitatively compared. This does not adapt the SIENA model in anyway, but rather suggests a specification that we seek to compare given a set of classifications on the context. The specification of model takes place in the defining mechanisms related to the three roots: endogenous network factors, individual factors and exogenous contextual factors. The specifications will differ given the knowledge of the nature of the social phenomena of interest (Shumate and Contractor, 2013). For example, if we consider cooperation in the car industry, different variables and/or parameters would be considered than, say, the

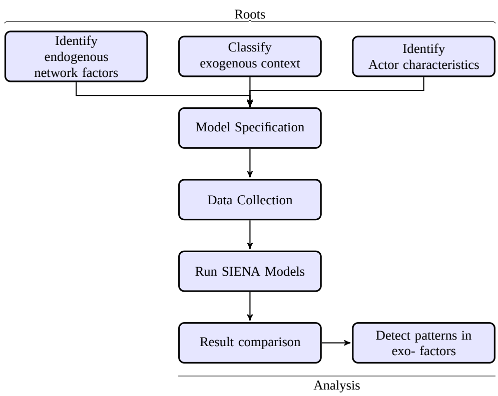



Figure 4.1: Method overview

logical relationships of building a car.

The first step is classifying a set of exogenous factors within a given set of interest. These contextual factors are unique to the population under study. We may think of them as organizational rules or policies that potentially influence individual networking decisions. They could be classified in any number of ways; in the specification below, we define them through interviews, but these contextual factors could also be specified and found, say, in a data store. Secondly, in proposing a recasting to compare models, we consider work in computer systems, which has developed a number of both qualitative and statistical guidelines for comparing simulations. These guidelines recommend keeping model differences low and comparing coefficients of effects as relative to the case and thus the research question (Goldsman and Nelson, 1998). If these models are identical in parameters and specifications then we are able to compare the role of different generative mechanisms with respect to these contexts.

We propose to compare these networks as we classify how a set of contextual conditions (as defined by the bounded network) relate to roles of significant generative mechanisms. This proposed methodology is a recasting of the current use of SIENA for investigating the contingent effect of exogenous contextual factors. Thus we emphasize the exploratory nature of this study, such that we do not seek to explicitly specify the models but rather seek to specify the strength of the SIENA models to provide a valid arena in which to explore methodological recastings of this nature in order to contribute to knowledge. In the following section we present the testing ground at which we test this method. Fig. 4.1 depicts the steps of our method.

## 4.5 Method

We delineate nine individual SIENA models based on the identical theoretical models to investigate the potential contingent effect of exogenous contextual factors. We explore these network dynamics in the setting of academic science. Additionally, conservative, theoretically identical models from Web data are identified and qualitatively compared.

### 4.5.1 Setting

Past decades have seen the emergence of new ways of working within all disciplines of scientific practice; these are evident from the growth in prevalence of an increasing number of co-authors on academic publications (Greene, 2007), to trends of working in teams in science (Falk-Krzesinski et al., 2010), to growing cross sector cooperation (e.g. triple-helix configurations (Van den Besselaar and Leydesdorff, 1996)) and shared laboratory pursuits (Shrum et al., 2007)). Given the perceived benefits of collaboration (Ding, 2011; Hudson, 1996; Sauer, 1988), scientific organizations, academic institutions and policymakers are seeking ways to facilitate and encourage collaboration through team science (Falk-Krzesinski et al., 2010), as well as through formal research initiatives such as grants (Defazio et al., 2009) and institutes (Falk-Krzesinski et al., 2010). Recent studies suggest that external factors influence the structures and tendencies towards collaboration (Gibbons et al., 1994). Additionally, institutional factors (Boardman and Corley, 2008; Ponomariov, 2007; Shrum et al., 2007) and requirements of funding schemes (Defazio et al., 2009) serve as evidence that moderators focus on the success of the network performance of individuals.

Computer science in The Netherlands is selected as a setting as it provides a number of academic institutions at which to examine different exogenous contextual factors within one national border. Computer science is selected due to the diversity and maturity of subfields, a tendency for collaboration, and a reliable online data. The following academic research universities are considered[1]: Universiteit van Amsterdam, Universiteit Utrecht, Technische Universiteit Delft, Vrije Universiteit, Technische Universiteit Eindhoven, Universiteit van Twente, Universiteit Leiden, Radboud Nijmegen Universiteit, Universiteit van Tilburg, and the Rijksuniversiteit Groningen. A source list of individual researchers was compiled from two sources:

(1) an official list of 434 tenured Dutch Computer science researchers in 2010 as acquired from The Nederlands Onderzoekdatabank, an official body that keeps records on research in The Netherlands,

(2) a manual Web query from the institutions as listed in the tenure list as post September 2011.

We look at the scientific collaboration networks of Dutch computer science researchers from 2006 – 2010.

### 4.5.2 Specification and Data Collection

**Network** In this study we investigate the practice of scientific collaboration networking via co-authorship networks using bibliometric data. Bibliometric data provides readily accessible, reliable, and scalable data. Bibliometric publication data is a representation of events. In this study, like the majority of studies on scientific collaboration which use bibliometric data, we project the data as relational network data to capture an assumed relationship on the part of the co-author at which to model the network states.

In particular we investigate conference proceeding publications. Conferences provide a number of clear timestamps discerning possible transition periods, as most conferences occur annually, with regular cycles. Additionally, within the

[1]There are a number of other universities in The Netherlands that also have computer science departments and Bachelor's training education, but they were not included in the Nationale Informaticakamer's review due to their lack of international research status. Thus the case selection is replicated in order to implement a secondary depiction of the cases.

field of computer science, conference proceedings are produced with little lag time.

To identify a valid and reliable set of co-authorship publication data for Dutch Computer Science researchers, the DBLP DataBase was queried. (DBLP is one of the most comprehensive bibliographic indices for the field of computer science.) All publications from our source lists were queried from 2006 - 2010. This list was manually cleaned to disambiguate names. From this list the name of the publication was queried to identify the unique author IDs of each author per publication. These unique author IDs were queried to pull full publication lists of each author (Dutch scientists and their coauthors). Conference proceedings are denoted in this data set by the BibTeX entry `@inproceedings`, allowing us to further query for proceedings-only publications. This yielded 9459 conference proceedings among the nine Dutch Computer Science departments. (Note: We removed single authored papers which we 6.5% of the conference proceedings (607 proceedings), due to the modeling of relations.)[m]

Given that we are exploring the effects of exogenous factors on individual networking behaviors, we have only included active researchers- those that were present in four of the five years within the data set. This ensures that the researchers have been embedded a significant period of time to limit potential other uncontrolled effects on their networking behavior. This resulted in 3639 Dutch authors and their co-authors. The publication data is then projected thus: Nodes are represented by individuals researchers and ties represent a shared publication. From this list we identified the collaborations and affiliated them back to the nine academic departments so as to create nine bounded networks of network panel data.

**Exogenous contextual factors** To further specify the effects of exogenous contextual factors on network emergence, we considered policies of the Dutch Computer Science departments as a measure of the conditions. We draw on findings from 25 semi-structured interviews of two experts and the head of the department (in all cases but one) from each of the nine Dutch academic departments to identify and later classify these types of exogenous contextual factors influencing scientific collaboration via policies. Expert interviewees were identified by two Dutch computer science researchers in two different sub-fields, both active in The Netherlands for over 20 years. From the list of nine academic departments, the two experts made a selection of three computer scientists from the criteria: those who could accurately reflect on organizational processes within the department. These were then crossed-checked for agreement towards the selection of two interviewees. Qualitative interviews provide a tool for gaining detailed descriptions of conditions, integrating multiple perspectives and thus bridging inter-subjectivities from multiple parties into a coherent story and describing a process, as well as developing holistic descriptions (Weiss, 1995).

[m] same dataset in (Birkholz et al., 2012)

Table 4.2: Policies at Academic Departments

| Academic department \ Policy | Professional Tenure system | Target Publication list | Incentives |
|---|---|---|---|
| $A$ | ✓ | | |
| $B$ | | | |
| $C$ | ✓ | ✓ | |
| $D$ | ✓ | | |
| $E$ | | ✓ | ✓ |
| $F$ | ✓ | | ✓ |
| $G$ | ✓ | | |
| $H$ | ✓ | ✓ | ✓ |
| $I$ | ✓ | | |

Interviews were guided by a set of structured questions that asked about policies that influence scientific collaboration. The interviews had an average duration of 60 minutes and were conducted in the interviewees university offices by the first author.

Findings from the interviews were compared based on a process of content analysis in order to assert a specific policy of the department. Those applicable included:

(1) the existence of a formalized professional tenure system that has a set of required specifications for advancing within the department,

(2) a target list of publication outlets, and

(3) any incentives – financial or time (i.e., more research time in exchange for grant acquisition, high impact journal publication and so forth).

The policies allow us to identify commonalities in department types, the level at which we compare the nine SIENA models. These are visualized in Table 4.2. For purposes of anonymity the departments are simply represented as Department $A$ - $I$.

**Individual characteristics** From the publication data set we computed data on a number of individual characteristics. Three characteristics are considered in the model: career age as a measure of expertise, number of co-authors, and number of previous conference proceedings as a marker of visibility. Studies on scientific collaboration have found a tendency for higher-tenured and -ranking researchers to collaborate (deB Beaver and Rosen, 1979). Consequently we include this in the model as noted by the first publication per author in the DBLP dataset. We also consider the researcher's access to socio-technical capital accounting for previous co-authorships, which suggests that access to potential coauthors in a field plays a key role in collaboration (Bozeman and Gaughan, 2007).

The number of coauthors per year per author is computed from the DBLP data set. Visibility of the researcher is also considered as the likely popularity through publication magnitude. These covariates are dynamic and computed per time period. Additionally, the models indicate a preference for homophily effects; thus, effects on expertise, visibility, and cosmopolitan are investigated.

We also consider institutional affiliation and nationality as homophily effects. For modeling purposes this covariate is constant, as we assume that all researchers currently conduct research at the institution identified (granted a small fraction of these researchers has likely changed institutions during this five year period). This decision is due to the limitations of the automatic collection of historical data on institutional affiliation. Each institution was identified through a query of two databases; for details on this process see (Birkholz et al., 2012).

**Endogenous Network Factors** We considered three network processes: transitivity, brokerage and popularity.[n] In this model we investigate the transitive ties effect, defined as the actors with whom the actor of interest $i$ is both directly and indirectly linked. Given the nature of the publication data used in this study, it is likely that groups of three or more individual researchers publish together, which could effect the strength of this effect in particular. Thus, we have selected to consider the most conservative measure of transitivity within SIENA, which is best used for undirected data and to control for this effect. Brokerage is defined by the ”betweenness” effect of the embeddedness in the network. It allows us to capture the situation of individual egos $i$ within their immediate networks. We also consider a degree popularity effect, which is defined by the sum of the degrees to the others to whom $i$ is tied. (Ripley et al., 2011)

**Model specification** Models in SIENA are specified by defining the rate and objective function of the theoretical aspects of the networks' interests that coevolve. The rate defines the speed of change within the network, and the objective function describes what this change entails. Considering the nature of data as an implied state from publication data to infer collaboration, we compile the network data into three time waves, based on years. We assume co-authorship has long decay, where contact remains for more than a year after the paper. Additionally, papers take time to produce, and, thus, it is not likely that a shared coauthorship emerges even yearly. Thus, we artificially control for this by combining years for the simulation (e.g. $T1$ is $2006{+}2007{+}2008$, $T2$ is $2007{+}2008{+}2009$, $T3$ is $2008 + 2009 + 2010$). Such a bridging is common practice in panel data. The objective function is made up of the first three effects which depend only on the network itself. The models were estimated under the standard options of SIENA (Snijders, 2005), with the exception of simulating 1000 runs instead of the standard 500 so as to overcome the nature of the data as inferred relations.

[n] A common mechanism includes reciprocity but given the nature of the data being non-reciprocal, we excluded this in our models.

Table 4.3: Descriptives

| | *A* | *B* | *C* | *D* | *E* | *F* | *G* | *H* | *I* |
|---|---|---|---|---|---|---|---|---|---|
| Active | 55 | 39 | 92 | 22 | 16 | 23 | 94 | 99 | 108 |
| Total | 713 | 603 | 962 | 274 | 210 | 272 | 1125 | 1599 | 1651 |
| # Coauthors | 1 - 99 | 1 - 81 | 1 - 125 | 1- 67 | 1 - 33 | 1 - 99 | 1 - 99 | 1 - 99 | 1-125 |
| | 5,21(9,18) | 4,59(8,22) | 4,94(8,99) | 5,32(8,48) | 4,27(6,44) | 5,99(10,29) | 4,48(7,70) | 4,49(8,36) | 3,77(6,81) |
| # Papers | 1 - 71 | 1 - 83 | 1 - 71 | 1 - 76 | 1 - 33 | 1 - 40 | 1 - 82 | 1 - 84 | 1-122 |
| | 3,74(7,26) | 3,42(7,43) | 4,59(7,60) | 3,32(6,69) | 3,18(5,07) | 2,81(4,84) | 3,88(3,33) | 3,33(6,87) | 3,52(7,89) |
| Scientific age | 1973-2010 | 1974-2010 | 1968-2010 | 1957-2010 | 1975-2010 | 1977-2010 | 1970-2010 | 1973-2010 | 1974-2010 |
| | 2003(6,10) | 2003(6,65) | 2003(6,47) | 2004(5,71) | 2002(7,97) | 2002(6,24) | 2004(5,71) | 2004(5,36) | 2004(5,62) |
| Nationalities per institute | 30 | 29 | 31 | 20 | 16 | 15 | 27 | 41 | 37 |
| # Institutes cooperated | 101 | 111 | 139 | 50 | 47 | 52 | 151 | 204 | 202 |

We then consider nine individual SIENA models that are identical in model development (e.g. with the same set of covariates); to this end, the models were run using RSIENA package 2.13.0. The experiments were performed on a DAS-4 site of VU University Amsterdam, one of the grid computer sites which belong to various academic institutions in The Netherlands.

We compare these findings given the classifications of the exogenous contextual factors within which the individual researchers are embedded. This allows us to indicate possible contingencies of specific exogenous contextual factors when considering the role of the roots in network emergence.

## 4.6 Results

Depictions of the collaboration networks are provided in the Appendix, Scientific Collaboration Networks of Dutch Computer Science departments 2006 - 2012. Descriptives of the academic departments and the coefficients are listed in Tables 4.3 and 4.4.

In particular, Table 4.3 shows the value of parameters for our SIENA models. The row "Active" represents the number of researchers from the corresponding academic departments, denoted by $A$-$I$, "Total" is the number of scientists from the institution and their coauthors, and "# Coauthors" (and "# Papers", or "Scientific Age") indicates the ranges of the number of coauthors (and papers, or the

Table 4.4: Coefficients

| | *Jaccard* | | *Degree* | | | *Ties* | | |
|---|---|---|---|---|---|---|---|---|
| | t1 > t2 | t2 > t3 | t1 | t2 | t3 | t1 | t2 | t3 |
| ***A*** | 0.61 | 0.63 | 2.58 | 2.68 | 2.37 | 921 | 954 | 844 |
| ***B*** | 0.56 | 0.65 | 1.84 | 2.32 | 2.54 | 562 | 700 | 767 |
| ***C*** | 0.67 | 1 | 5.1 | 5.49 | 5.49 | 2452 | 2642 | 2642 |
| ***D*** | 0.64 | 0.64 | 2.26 | 2.27 | 2.32 | 309 | 311 | 318 |
| ***E*** | 0.58 | 0.57 | 1.83 | 1.8 | 1.81 | 192 | 189 | 190 |
| ***F*** | 0.55 | 0.66 | 1.77 | 2.2 | 2.43 | 240 | 299 | 330 |
| ***G*** | 0.61 | 0.58 | 2.45 | 2.64 | 2.64 | 1376 | 1486 | 1483 |
| ***H*** | 0.59 | 0.58 | 2.43 | 2.74 | 3.1 | 1946 | 2190 | 2480 |
| ***I*** | 0.64 | 0.56 | 2.26 | 2.41 | 2.22 | 1864 | 1991 | 1831 |

first publication year) of the scientists from the institution (with average value and Std). The last two rows are the number of different nationalities represented by the number of academic institutions, and/or other affiliations of co-authors. Table 4.4 provides further descriptions of the models, including the Jaccard coefficient, change in degrees and ties between the time waves.

We analyzed two models per academic department. We first addressed endogenous factors (Table 4.5), and then included individual factors (Table 4.6) in a separate model for each of the nine academic departments. Significant effects in the models are signified by a * for $p$ values above 0.05. The tables also denote the contextual variables of the existence of a tenure system, publication list, and a combination of or lack thereof, to identify possible patterns in connection to the emergence of generative mechanisms which explain the network structure. Specific results of the individual academic departments are included here.

As we see from Table 4.6, the network of University $A$ has significant transitive ties ($\beta = 0.94$; $p < 0.01$), as well as the popularity to collaborate with an author with a high number of ties during this period ($\beta = 21.18$; $p < 0.01$). The addition of actor characteristics by way of covariates does not have an effect in explaining the networking decisions of the researchers. In University $B$ the network model shows that all three network effects influence the maintenance and emergence of ties. Researchers within this network have a significant tendency to make transitive ties ($\beta = 36.16$; $p < 0.01$). When considering covariates in the network, the evolution of the network can in part also be explained by researcher collaboration with a researcher who has a lower number of coauthors ($\beta = -2,56$; $p < 0.01$), as well as with researchers outside of their university

Table 4.5: Model 1 – Network effects

| | *A* | | | *B* | | | *C* | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Estimate** | **SE** | **Sig > 2** | **Estimate** | **SE** | **Sig > 2** | **Estimate** | **SE** | **Sig > 2** |
| Degree | -4,4 | 0 | -10.44* | -6,3771 | 0 | -24.07* | -3,2548 | 0 | -48.66* |
| Transitivity triads | 1,145 | 0 | 18.94* | 2,2454 | 0 | 36.16* | 0,2347 | 0 | 12.04* |
| Betweenness | -0,1987 | 0 | -0,478 | 1,6223 | 0 | 8.94* | -0,0651 | 0 | -2.50* |
| Degree of alter | 0,0593 | 0 | 21.18* | 0,0644 | 0 | 22.21* | 0,0471 | 0 | 19.63* |
| | | | | | | | | | |
| | *D* | | | *E* | | | *F* | | |
| | **Estimate** | **SE** | **Sig > 2** | **Estimate** | **SE** | **Sig > 2** | **Estimate** | **SE** | **Sig > 2** |
| Degree | -4,2 | 0 | -11.02* | -4,3994 | 0 | -11,66 | -3,4422 | 2 | -1,88 |
| Transitivity triads | 0,759 | 0 | 8.62* | 1,2433 | 0 | 14,56 | 1,1918 | 0 | 11.26* |
| Betweenness | -0,0145 | 0 | -0,053 | 0,171 | 0 | 0,57 | -0,9039 | 3 | -0,289 |
| Degree of alter | 0,0833 | 0 | 14.875* | 0,1364 | 0 | 5,33 | 0,1523 | 0 | 10.50* |
| | | | | | | | | | |
| | *G* | | | *H* | | | *I* | | |
| | **Estimate** | **SE** | **Sig > 2** | **Estimate** | **SE** | **Sig > 2** | **Estimate** | **SE** | **Sig > 2** |
| Degree | -4,946 | 0 | -18.94* | -3,9893 | 0 | -39.81* | -4,3393 | 0 | -23.37* |
| Transitivity triads | 1,4902 | 0 | 21.35* | 1,1156 | 0 | 46.87* | 1,1197 | 0 | 34.88* |
| Betweenness | 0,1345 | 0 | 1,24 | -0,1759 | 0 | -2.67* | -0,2786 | 0 | -1,44 |
| Degree of alter | 0,0664 | 0 | 18.97* | 0,0489 | 0 | 27.17* | 0,0417 | 0 | 27.8* |
| | | | | | | | | | |
| | | Both a formalized tenure system & publication list | | | | | | | |
| | | formalized tenure system | | | | | | | |
| | | publication list | | | | | | | |

Table 4.6: Model 2 – Network and Covariate effects

| | A | | | B | | | C | | |
|---|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | Sig > 2 | Estimate | SE | Sig > 2 | Estimate | SE | Sig > 2 |
| Degree | -3,9188 | 16 | -0,24 | -3,7874 | 0,9 | -4.22* | -2,7189 | 0,13 | -21.08* |
| Transitivity triads | 1,2341 | 1,2 | 1,03 | 1,3894 | 0,15 | 9.12* | 0,2986 | 0,04 | 8.44* |
| Betweenness | -0,5788 | 24,8 | -0,02 | -0,2125 | 0,89 | -0,24 | -0,2432 | 0,06 | -4.24* |
| Degree of alter | 0,1167 | 0,58 | 0,2 | 0,0502 | 0,02 | 2.51* | 0,0589 | 0 | 15.10* |
| Institute | -0,1534 | 3,75 | -0,04 | -1,7437 | 0,3 | -5.79* | -0,0604 | 0,16 | -0,37 |
| Age | -0,551 | 6,48 | -0,08 | -0,3902 | 0,39 | -1,002 | -1,1935 | 0,13 | -8.93* |
| Nationality | -0,6056 | 2,95 | -0,2 | 0,6015 | 0,31 | 1,93 | 0,4503 | 0,13 | 3.53* |
| Paper | -0,0757 | 0,61 | -0,12 | 0,0324 | 0,02 | 1,41 | -0,0473 | 0,01 | -7.51* |
| Coauthors | 0,0194 | 0,12 | 0,16 | -0,0386 | 0,02 | -2.56* | 0,0117 | 0,01 | 2.21* |
| | | | | | | | | | |
| | D | | | E | | | F | | |
| | Estimate | SE | Sig > 2 | Estimate | SE | Sig > 2 | Estimate | SE | Sig > 2 |
| Degree | -3,7638 | 18,7 | -0,2 | -4,7244 | 2,45 | -1,93 | -4,4853 | 10,1 | -0,445 |
| Transitivity triads | 0,9941 | 1,78 | 0,56 | 1,3395 | 0,52 | 2,5995 | 1,2085 | 2,38 | 0,51 |
| Betweenness | -0,6641 | 48,1 | -0,01 | 0,445 | 1,82 | 0,2446 | 0,3115 | 5,86 | 0,05 |
| Degree of alter | 0,0189 | 0,38 | 0,05 | 0,1586 | 0,1 | 1,5309 | 0,14 | 0,33 | 0,42 |
| Institute | -3,0766 | 48,9 | -0,06 | -0,7528 | 0,82 | -0,92 | 0,0931 | 3,56 | 0,03 |
| Age | 0,9131 | 3,04 | 0,3 | -1,4866 | 0,77 | -1,927 | -0,4377 | 2,81 | -0,16 |
| Nationality | 0,473 | 6,41 | 0,07 | 0,4642 | 0,46 | 1,0173 | -0,0836 | 1,08 | -0,078 |
| Paper | 0,0534 | 0,32 | 0,16 | -0,0241 | 0,11 | -0,217 | 0,0187 | 0,19 | 0,1 |
| Coauthors | 0,0309 | 0,33 | 0,09 | -0,0072 | 0,06 | -0,121 | -0,0623 | 0,12 | -0,5 |
| | | | | | | | | | |
| | G | | | H | | | I | | |
| | Estimate | SE | Sig > 2 | Estimate | SE | Sig > 2 | Estimate | SE | Sig > 2 |
| Degree | -4,8542 | 0,95 | -5.11* | -3,7387 | 0,1 | -37.69* | -4,5119 | 0,12 | -36.59* |
| Transitivity triads | 1,5938 | 0,25 | 6.36* | 1,1616 | 0,02 | 51.86* | 1,1439 | 0,03 | 41.75* |
| Betweenness | 0,1899 | 0,4 | 0,48 | -0,2407 | 0,09 | -2.83* | -0,0733 | 0,11 | -0,65 |
| Degree of alter | 0,0855 | 0,02 | 3.46* | 0,0448 | 0,01 | 5.33* | 0,0465 | 0,01 | 3.39* |
| Institute | -0,498 | 0,32 | -1,56 | 0,3913 | 0,12 | 3.22* | -1,1776 | 0,17 | -6.90* |
| Age | -0,846 | 0,23 | -3.67* | -1,3054 | 0,11 | -12.08* | -1,1919 | 0,16 | -7.46* |
| Nationality | -0,2666 | 0,31 | -0,86 | -0,0334 | 0,07 | -0,468 | 0,2782 | 0,11 | 2.595* |
| Paper | -0,033 | 0,02 | -1,43 | -0,0103 | 0,01 | -1,157 | 0,0268 | 0,01 | 2.16* |
| Coauthors | 0,0094 | 0,01 | 1,175 | 0,0171 | 0 | 5.18* | -0,0402 | 0 | -9.35* |
| | | | | | | | | | |
| | | Both a formalized tenure system & publication list | | | | | | | |
| | | formalized tenure system | | | | | | | |
| | | publication list | | | | | | | |

institution ($\beta = -5,79$; $p < 0.01$). At University $C$, the network effects play a significant role in explaining network evolution; all remain significant when adding actor characteristics to explain dynamics. The network of University $D$ has significant tendencies for transitivity $\beta = 8.62$; $p < 0.01$), as well as the popularity of connecting with alters ($\beta = 14.88$; $p < .01$).

Consideration of covariates does not significantly influence the network effects. In University $E$, again, there is a tendency for making transitive ties ($\beta = 14.56$; $p < 0.01$); there also exists a popularity effect ($\beta = 5.33$; $p < .01$). Tendency for transitivity ($\beta = 2.60$; $p < 0.01$) remains significant when considering other covariates in the second model. In University $F$, the network tendencies in the first model confirm a tendency to make transitive ties ($\beta = 11.26$; $p < 0.01$), as well as the role of popularity ($\beta = 10.50$; $p < 0.01$). In the second model the effects lose their significance. At University $G$, tendencies for transitive ties ($\beta = 21.35$; $p < 0.01$) and popularity ($\beta = 18,97$; $p < 0.01$) are of significant influence. Again, at University $I$ transitive ties ($\beta = 34.88$; $p < 0.01$) and popularity ($\beta = 27.80$; $p < 0.01$) have an effect on network dynamics. In the second model, these two increase in strength.

Thus, our results show that researchers in all of the academic departments have a tendency to initiate and maintain collaboration ties that are transitive. In addition there is a tendency to initiate and maintain collaboration ties with popular authors, which lead to the so-called Matthew effect (Merton, 1968) of preferential attachment (Albert and Barabási, 2002), where researchers with high numbers of collaborations increase exponentially. Tendencies to connect with actors with high betweenness (those who hold a brokering position and have, theoretically, greater potential to control communication and information within the network (Burt, 2005)) vary in both significance and direction of effect between the departments. Within all academic departments, we find that researchers have distinct tendencies to connect with actors, and the strengths of these effects and existent of the effects vary between the departments. In the second model we add actor characteristics as an explanation for network emergence. Results show even greater distinctions between the models, wherein some model effects are not significant, while others increase in strength.

In comparing the models, we aim to explore potential differences in effects that can be explained through exogenous contextual factors. First, looking at Model 1, tendencies of effects are similar in departments C and H, with both displaying a positive and significant tendency for transitive ties, as well as a negative likelihood to collaborate with researchers that occupied broker positions. These effects remain significant when considering covariates, with the distinction that Department H has a tendency to work internally with other colleagues, instead of with those outside of the department. Both departments have formalized tenure systems and a list of target publications. These two are distinctly different from the significant effects of Department I, which only has a formalized tenure system. In this case, a tendency to work with Dutch researchers – as well as those with high visibility in previous publications – contributes to the generative mechanisms within this network. In departments with just one policy (departments $A$, $D$, $G$, $E$, and $F$), a variety of effects has resulted in not one unique pattern can be attributed to one particular policy. Departments $A$, $D$, $E$, and $F$ in particular lose significant effects when considering covariates which relate to the combination of individual and endogenous roots. Department $B$ and $G$ do not differ greatly when covariates are tested, though $B$ is lacking policies.

## 4.7 Conclusion

This research proposed an analysis of the effects of organizations on steering individuals' networking behaviors through the implementation of a recasting of the use of SIENA models. We investigated the potential effects of exogenous contextual factors as viewed through the effects of the policies of Dutch Computer Science departments on the generative mechanisms that explain the network structures of research collaborations via co-authorship. Such a method allowed us to consider context as a determinant of structures, which was done through the qualitative comparison of models retaining the methodological strength of the commonly accepted method of SIENA through a recasting of its formal use. Such a method allows us to identify potential patterns for exploring the modi operandi of social networks.

Findings showed that the existence and strengths of effects varied among departments. Linking these distinctions back to the specific classifications of exogenous contextual factors, we find patterns in the specific factors in particular departments with clear policies for attempting to steer outputs by way of a formalized tenure system and a publication list which generated networks with tendencies for transitive ties, as well as a negative likelihood to collaborate with researchers that occupied broker positions. Thus, a context with focused policies, such as tenure and a target list, displays network structures where researchers work together in close collaboration. This coincides with findings on innovation that suggest that dense networks, with low barriers between actors (by way of relations), facilitate information sharing (Coleman, 1988; Granovetter, 1985). Departments with one or no policies indicated no root, individual or endogenous, can be attributed to the emergence of these network structures, suggesting the influence of a variety of network behaviors on the part of individual researchers in navigating collaboration. Context partly explains the local social mechanisms that researchers undertake to collaborate, and thus researchers should consider garnering awareness of both the structures of their collaboration networks and the policies that aim to shape their publication outcomes and, thus, collaboration decisions. Future work should consider a larger set of cases to greater specify the role of the three roots in network structures. Where exogenous contextual factors are a determinant to invoking roots, different factors that influence the emergence of distinct mechanisms should be considered.

These network models implicitly consider the role of an exogenous contextual factor. Specifically, the algorithms used to model dynamics originating in the natural sciences conceptualize dynamics of entities (e.g., atoms, molecules, cells and the like) having the capacity to interact with other entities, where their behaviors are constrained by the conditions of the environment or physical space in which they reside. These conditions steer the trajectories of individual entities, thus giving rise to a (fixed) set of interaction options among the known possible interactions between two entities. Consequently, dynamic network models implicitly assume that context is a contingency to the emergence of different network structures. However, the large majority of these network models, also used in the Social Sciences, include theoretical models developed to investigate social network structures and ignore context as an explanatory variable through considering small or boundary-driven networks.

This recasting of an established method such as SIENA (in this chapter) provides a unique development to a researcher's toolbox. It opens a new testing ground for the exploration of the contingencies of context. It aids in identifying the most typical explanations for the networking behaviors observed in these different contexts. Although this provides additional insights for exploring possible mechanisms, the story remains incomplete. Even in Snijders and Baerveldt (2003)s multilevel network approach in SIENA, which yields typical networking tendencies given a sample of a population considering sequentially within and between-network analysis, the assumption remains that we identify individual networks by an implicit boundary  a classroom, a school, an organization. Thus, although this model would allow us to test whether context plays a role in the emergence of relationships, it fails to capture or identify how context potentially influences the emergence of different generative mechanisms that could explain differences in structures. Additionally, it fails to consider inter-network interactions, when in reality these networks may overlap. Considering networks where there are known overlaps, a more valid approach would be to undertake a meta approach that considers all the behaviors of these individuals, as constrained by the rules of the context, but that also have the potential to interact within a field of others that are embedded in different contexts. In considering a meta-model of this sort, we are able to observe how context serves as an effect on the emergence of network structures. In an attempt to respond to this call and depict a more complete and valid understanding of context as an effect the emergence of network structures I experiment with the use of a mean-field model in the following chapter.

# 5 Scalable Analysis for Large Social Networks: The data-aware mean-field approach[o]

## 5.1 Abstract

Studies on social networks have proved that endogenous and exogenous factors influence dynamics. Two streams of modeling exist on explaining the dynamics of social networks:

1) models predicting links through network properties, and

2) models considering the effects of social attributes.

In this interdisciplinary study we work to overcome a number of computational limitations within these current models. We employ a *mean-field model* which allows for the construction of a population-specific model informed from empirical research for predicting links from both network and social properties in large social networks.. The model is tested on a population of conference coauthorship behavior, considering a number of parameters from available Web data. We address how large social networks can be modeled preserving both network and social parameters. We prove that the mean-field model, using a data-aware approach, allows us to overcome computational burdens and thus scalability issues in modeling large social networks in terms of both network and social parameters. Additionally, we confirm that large social networks evolve through both network and social-selection decisions; asserting that the dynamics

[o]This paper is published in the form presented here as (Birkholz et al., 2012). Thus it is written in the plural we form.

of networks cannot singly be studied from a single perspective but must consider effects of social parameters.

## 5.2 Introduction

Dynamics of social networks are receiving increasing attention in multiple research domains (Ahuja et al., 2012; Albert and Barabási, 2002; Snijders et al., 2010). Theoretical developments posit that dynamics are influenced by network (Barabási and Albert, 1999) and social processes (Snijders et al., 2010); with recent theory suggesting that the two co-evolve (Ahuja et al., 2012). Methods to explore dynamics of networks traditionally implement evolving graph models, using inferential statistics to assert the likelihoods of the creation, maintenance or dissolution of edges. Two distinct classes of modeling exist:

(1) exclusively modeling the effect of network structures on dynamics (Liben-Nowell and Kleinberg, 2007; Moore et al., 2006), and

(2) modeling effects of social parameters and network effects for small networks ($\sim$ 1000 nodes) (Snijders et al., 2010).

Both types of models prove that network processes affect the dynamics of networks. Network models have been able to accurately predict a small percentage of edges, suggesting that dynamics may also be fed by other processes. Social-parameter models have proved social attributes, in combination with network structures, play a role in network dynamics.

Despite this growing knowledge from both model classes, these models have limitations. The main limitation relates to using an evolving graph model which calculates statistical probabilities of individual nodes. This approach generally leads to a super-linear growth in computational load as the network size increases, partly caused by the quadratic growth in the number of links that need to be considered. Both models attempt to overcome this through different means. One is limited to either testing the effect of a few parameters on a large network, or a number of parameters on small networks. Consequently, neither provide a terrain to empirically confirm the effect of both network and social parameters in large social networks.

In order to better understand the dynamics of large social networks, a different computational approach must be taken to overcome the issue of scalability in present models. In this paper we review the two existing model classes used to investigate dynamic social networks, and present a model for overcoming a number of acknowledged limitations. Using a mean-field model approach we are able to overcome scalability issues in previous models through aggregation of individual nodes. Parameters are developed using a data-aware approach which combines empirical research from Social Science and standard inferential statistics to develop a population-specific model for exploring the dynamics of collaboration in science.

We consider the question whether mean-field modelling allows us to describe the behavior of a social system, considering a number of network and social parameters. In this first application of the mean-field model to large social networks, we aim to explain the effect of a set of parameters governing networking patterns of collaboration in Dutch Computer Science (CS). Four parameters are considered in this research: institutional affiliation, scientific age, cosmopolitanism of knowledge production, and visibility of the scientists. We prove that mean-field models expand the empirical testing ground of dynamic network models through increased scalability. This allows us to better understand dynamics of large social networks, covering space that has not been investigated in the past using a mean-field approach.

The paper is set up as follows. In Section 5.3 we review the state of social network models, specifically highlighting the limitations of present models. In Section 5.4 we explain the mean-field model, discussing in detail the computational advantages of the model as well as the steps taken to implement a data-aware approach for improved specifications. In Section 5.5, we test the model on the coauthorship networks of papers from the conference proceedings for Dutch computer scientists, collected from the DBLP data set for 2006 – 2010. Finally, we conclude with the results and implications for scalable, data-aware modeling solutions for explaining dynamics of social networks.

## 5.3 Network Models

The evolution of a network is driven by the addition, maintenance, and dissolution of interactions (edges) between nodes over time. Evolving graph models are the most commonly implemented models to explain the dynamics of networks (Barabási et al., 2002; Grossman, 2005; Newman, 2004). These models assume that nodes are added one-by-one to the network, in discrete time. They infer the probability of a link emerging given a node-transition rate using a Markovian model of simulation. Within this model type two distinct approaches exist investigating social network dynamics:

(1) global network-structure link-prediction models, and

(2) social-parameter models integrating social factors into link prediction.

Models with pure network-structure prediction assumptions derive from the vast research on global network structures. Studies on network properties confirm that many real-world networks display small-world properties in which high node clustering is combined with short average internode distances (Newman, 2004; Watts and Strogatz, 1998). Networks have also been found to behave according to a power-law scale-free phenomenon where a relatively small number of nodes have numerous connections (Akkermans, 2012; Albert and Barabási, 2002; de Solla Price, 1965). Additionally, networks have properties of clustering hierarchies (Albert and Barabási, 2002), and tendencies of transitivity or "triangles of interaction" describing the manner in which ties between node $A$ and $B$, and between node $B$ and $C$ facilitate a likely tie between $A$ and $C$.

From this knowledge on network properties a second generation of studies emerged addressing how a social network can be modeled using properties intrinsic to the network. These global network-structure link-prediction models provide insight into not yet identified or observed linkages (Krebs, 2002), as well as to infer not directly observed likely links (Goldberg and Roth, 2003; Popescul and Ungar, 2003; Taskar et al., 2003). Within these studies two approaches are taken to predict links:

(1) computing node-level measures from greater network structures and,

(2) meta-level analyses.

In this study we consider only node-level measures (which are comparable to the gap we aim to fill in this research), while still maintaining the network structure.

Several approaches for predicting social network linkages have been proposed, for a complete list see (Liben-Nowell and Kleinberg, 2007). Despite the extensive research of different measures used to model the network dynamics, all of these models suffer from low fitness, with random link prediction performing just as well as Katz's model of path collection- predicting links by the sum of collected path lengths per individual (Katz, 1953). This has led informaticians to explore the effects of additional parameters in understanding network dynamics. A second model type works to address the effect(s) of social parameters on the dynamics of social networks. The justification for these models arose from research on social networks which proved that social selection plays a key role in relation formation (Ennett and Bauman, 1994; Granovetter, 1973; Krackhardt, 1992). Models of this type allow us to question how a social network can be modeled using both network and social properties of nodes. These models also infer edges through evolving graph models but consider state spaces with both network and social parameters. Two model types are commonly used to investigate the inference of these dual parameters: stochastic actor models (SIENA)(Snijders et al., 2010) and exponential random graph models (ERGM) (Robins et al., 2007).

The key distinction in these models, from the network-only models, is the combination of link prediction based on both local effects, as well as on "social circuits" that capture the influence of more distant ties on behavior (Robins et al., 2007). This leads to an exponential growth of the state space due to the consideration of more parameters, requiring extensive computing power in prediction. Given the computational complexity of calculating this for every node these models are not easy to develop in a way that convergence emerges in large networks (Robins et al., 2007). Consequently, these classes often limit the size of networks through a theoretical boundary of inferring statistics for a bounded network. This reduces the burden of having to perform computations on potentially very large graphs, but also effectively limits application to small networks ($\sim 1000$ nodes).

In summary, these two model classes provide a testing ground to explore dynamics, but are both not without limitations. Both network and social parameters have scalability problems. As we discuss next, in order to empirically explore the effect of both network and social parameters on large social network dynamics a scalable solution is required.

## 5.4 Modeling Framework

We propose a *mean-field approach* for studying social networks; (equally behaving) individual nodes are grouped according to their *states*. This approach is used for an optimized analysis of large-scale systems, allowing for a prediction of the average behavior of the system. The mean-field theory has been applied previously, e.g., to large-scale gossip systems in (Bakhshi et al., 2010). Concisely, the state of the system is represented by a distribution, or a vector of fractions of nodes $\delta_s(t)$ in each state $s$ at time unit $t$. The evolution of the stochastic system is governed by a so-called master equation of the form:

$$\delta(t+1) = M_{\delta(t)} \cdot \delta(t) \tag{5.1}$$

$M_{\delta(t)}$ is the matrix, each entry of which is a transition probability from a state $s$ at time $t$ to state $s'$ at time $t+1$. Thus, we are effectively reducing the global state space, thereby increasing the computational efficiency of the model, and in turn, allowing us to consider more parameters as well as more nodes.

Moreover, we use the notion of *classes*, introduced in (Bakhshi et al., 2010), to distinguish between equally behaving nodes affiliated to different categories. To this end, the mean-field model predicts average behavior of sets of nodes of each class given a number of social and network parameters. We highlight the modelling steps:

**Forming a model** In order to model the network, first we need to define the system in the form of its parameters. This will form a state of the system. Given the type of network under study, the effects of system parameters are considered using either manual classification or statistical classification (e.g., (Bishop, 1995)) to identify the set of significant parameters to form states and classes. For example, some parameter $u$ can be a theoretically informed organizational constraint (e.g. an organization, a background, etc).

**Applying abstraction refinement** The theory underlying the mean-field model requires also the population of each state to be large enough to be approximated by the law of large numbers. The size of the population in a sampled data set may force one to consider further abstraction for the ranges of the parameters, thereby reducing the size of the system state space. For instance, if chosen parameters for the system are the number of papers per author $p \in \mathbb{N}$ and the number of an author's coauthors $c \in \mathbb{N}$, the number of possible states of the system will simply be a product $\mathbb{N} \times \mathbb{N}$. Some parameters can be restricted in their value ranges without loss of the accuracy of the model itself.

**Computing the model input** To execute the model, input data is needed on the initial state of the system, as well as on distributions for networking behavior, which will be used for the matrix $M_{\delta(t)}$. The input distributions for the mean-field model include three categories:

(1) communication: the interaction between nodes

(2) idle: a state of no interaction, and

(3) collision: the disappearance or decay of an interaction.

The distributions of interaction (links, from a graph-theoretical perspective) are estimated for each class, which determines the nonuniform behavior by different classes for the model. We compute these distributions statistically from the sampled data set.

**Estimation of distributions** The aforementioned transition probability distributions are determined using a discrete-time model to identify the optimal time slicing for the studied data set. Such a time slice corresponds to one time unit in the model. The distribution for probability of transition from one class to another one is also used in the master equation (5.1) (for a more detailed equation, cf. (Bakhshi et al., 2010, Fig. 10). The method used for estimation of the probability distributions is a Hidden Markov Model (HMM) (Rabiner, 1990).

**Applying automated mean-field framework** Armed with the knowledge regarding states, classes and transition rates, obtained from the previous steps, we apply an automated mean-field framework to infer average behavior of the system. We repeat the earlier steps until all parameters are included for a time period covered by the data set. We use the resulting mean-field model to make average link predictions on the system given the parameters under consideration. The model provides a number of advantages over models discussed in Section 5.3, such as greater flexibility in modeling behavior of nodes through a number of processes. The use of HMMs provides an additional round of probability in node interactions, to compensate for the aggregation. Moreover, such a model allows us to consider both social parameters as well as network structures. Unlike simulation or deployed models, the model is flexible given a theoretical knowledge of the interactions under study. In analyzing the system under question we set the formal specifications which provide detailed processes of specification.

**Considerations for extensions of social networks** The challenge in applying the mean-field model to social networks is to derive accurate predictions of the local behavior of the nodes within defined classes. Particularly, for social networks, model abstractions need to be done using a data-aware approach. A data-aware approach implies that both classes and parameters are informed through an intense, robust knowledge of the system under study, as well as the content of edges in the network data. It is a requirement that this is approachable through a theoretically or empirically grounded conceptual scheme on both the system under study and the mechanisms that inform the parameters considered in simulation models. Consequently, not all social networks and or systems can be analyzed using such an approach.

Additionally, we argue for an interdisciplinary approach in development of the model as data needs to be intensely explored to inform parameters by both a data engineer and validated by social scientists or informed experts of the system under study. This implies, unlike other models, that the data-aware approach is essential to determining accurate results, which can be compared in model-fit tests. This results in a model that specifically fits the needs of the system under study, and which can be adapted per population given the basic set of rules for abstraction we describe. In the next section we lay out the general steps for the application of a mean-field model.

## 5.5 Application

As discussed in the previous section a set of requirements are necessary for implementing a mean-field model to investigate the effect of social and network factors on network dynamics: network data, parameter data, and knowledge from empirical studies of the system under study. We explain the case studied here and detail the abstraction steps undertaken to model the effect of network and social parameters on network dynamics.

### 5.5.1 Network data

A majority of computational analyses of large social networks implement coauthor or similar co-occurrence networks to examine network dynamics (Albert and Barabási, 2002). Coauthorship networks, via publication data, provide a representation of a specific social interaction- successful collaboration, in producing an output- dissemination of knowledge through publication. Moreover, publication data is readily accessible on the Web providing large, reliable, and scalable data sets to model network dynamics.

In addition to the use of coauthorship data to study network dynamics, empirical studies on coauthorship provide a framework to develop measures to consider in the model testing. In science studies, coauthorship is a standard measure for collaboration in science. Collaboration is increasingly common in science;

from the near disappearance of single-authored papers to the growth in prevalence of an increasing numbers of coauthors on academic publications (Greene, 2007). A decade of studies on collaboration in science have proved the effect of different social variables on collaborative behavior of scientists (Bozeman and Corley, 2004; Stokols et al., 2008b). Recent studies have found that task types and a number of external factors influence collaborative behavior of scientific processes (Stokols et al., 2008b). Both institutional and short geographical distances play a key role in the collaborative behavior of scientists (Rodriguez and Pepe, 2008; Uzzi, 2008). Given these studies we have a basis at which to both test informed parameters and link findings to knowledge on collaborative tendencies of scientists.

In this paper we explore a system of collaborative behavior of scientists in testing the mean-field model for large social networks. We select one nation and discipline – Dutch computer scientists, to investigate dynamics as to limit known exogenous effects of different knowledge production practices between disciplines and nations. Effectively, we comment only on the average behavior of the system of Dutch CS. The field of CS was chosen for three reasons: the traditions of the field with a diversity of subfields within the discipline; the known tendency for collaboration through coauthorship; the validity and reliability of online sources documenting publications. The Dutch context provides a diversity of cases at which to examine different institutional processes.

A source list of 434 tenured Dutch computer scientists in 2010 was acquired from the Nederlands Onderzoekdatabank, an official body that keeps records on research in the Netherlands. To identify a valid and reliable set of coauthorship data for the Dutch computer scientists a snapshot of DBLP DataBase was queried. (DBLP is one of the most comprehensive bibliographic indices for the field of CS.) Within this set the list of Dutch computer scientists was queried for all publications of scientists from 2006 - 2010 (the year of our list of tenured scientists). This list was manually cleaned to disambiguate names. From this list the name of the publication was queried to identify the unique author IDs of each author per publication. These unique author IDs were queried to pull full publication lists of each author (Dutch scientists and their coauthors).

Conference proceedings were selected for the case study as conferences in CS require at least one author to physically present work at a conference to be published. Conferences provide a good fit for the assumption of interaction in previous computer models as a potential meeting points for coauthors. Additionally, it provides a number of clear timestamps discerning possible transition periods, with most conferences occurring annually, with regular cycles. Conference proceedings are denoted in this data set by the BibTeX entry `@inproceedings`, allowing us to further query for proceedings-only publications. This resulted in 3639 scientists, and 2757 conference-proceeding publications. Nodes represent individual scientists and links represent shared coauthorship of proceedings. From this data set of individual authors we also collect data on the social parameters.

### 5.5.2 Parameters

In this study we aim to include parameters that are informed from previous empirical studies in the field of science studies. Four parameters are considered in the model: scientific age, cosmopolitanism of knowledge production, visibility, and institutional affiliation. For the collection of social parameter data in this study the Web is used, providing a reliable method for collecting meta-data on scientists within publication records (Mika et al., 2006). The use of Web data as the source of meta data is integral in this first model development as it reduces the burden of data collection of social variables (compared to traditional social science data of surveys or interviews). This allows us to quickly test the effect of social parameters on behavior with a considerable amount of reliability from merging meta-data from additional online databases.

The parameters – scientific age, cosmopolitanism of knowledge production, and visibility are calculated from within the DBLP data set. Scientific age was selected because tenure and rank are both said to play a role in collaborative behavior of scientists, with scientists of a higher tenure more likely to collaborate than mid-range, tenure-seeking colleagues (deB Beaver and Rosen, 1979). We first noted publication per author in the DBLP data set for which we compute per year per author as his or her scientific age. A second parameter, cosmopolitanism, relates to the socio-technical acquired capabilities of scientists suggesting that access to potential coauthors in a field plays a key role in collaboration (Bozeman and Corley, 2004). This parameter was measured through previous coauthorship experience. The number of coauthors per year per author is computed from the DBLP. The third parameter aims to comment on the visibility of the scientist. The visibility of the scientist is the likely popularity through publication magnitude. These three parameters allow us to consider a number of possible social factors that are not network effects but rather social attributes on the scientists' networking behavior.

One additional parameter was collected for consideration in the model – the institution. Previous studies proved that the institution is statistically significant with respect to how scientists collaborate (Rodriguez and Pepe, 2008; Uzzi, 2008). The institution is identified through a query of two databases. These data are considered static in this model, unlike the previously mentioned data, as we assume minimal change of institution in the five-year period under study. The automatic collection of historical data on institutional affiliation is not currently stored in one database, to our knowledge, thus we assume a five-year period as a valid period of time to accurately measure inference. A query using Microsoft Academic Search – a database which includes the DBLP data set is used to identify institutions. To locate additional missing data another database, ArnetMiner.org was used. The remaining unidentified institutions were queried manually giving us a total of 1358 identified institutions. In order to disambiguate institutional names, to have a reliable and valid set of data, this list was queried in geocoding Web service Yahoo! PlaceFinder (Yahoo! PlaceFinder, 2012). This query provides a proximity measure for each institution and a uniform institutional affiliation based on common GPS coordinates.

These four parameters provide a setting to explore the application of the mean-field model in large social networks. The occupancy measure at time $\delta(t)$ in our model is the fraction of people in state $(p, c, h, u)$, where $p$ is a number of publications, $c$ is a number of coauthors, $h$ is scientific age, and $u$ is affiliation. We test the following social science hypothesis: institutions effect the patterns of collaborative behavior (by behavior we mean average number of coauthors, and average number of papers). In addition to these social parameters we also include the network parameter of transitivity. As discussed in section 1, social networks have tendencies of transitivity (Albert and Barabási, 2002; Newman, 2004). We consider the social parameters in predicting the triadic interactions between nodes.

### 5.5.3 Classes abstraction

In principle, any of our parameters could be considered a class. When studying a social system, however, we need to consider known social and organizational constraints. In order to define a class we investigate the four possible parameters under consideration in this model. We first consider known effects.

Our system is already bounded by the selection of one national science structure and one scientific discipline. The effect of the institution provides a valid and logical boundary at which to explore aggregation. Additionally, we know that geographical location also plays a key role in collaboration, which we aim to consider in the abstraction. Consequently, we employ institutions as classes in our mean-field model, and as one of the parameters $u$ contributing to a state $(p, c, h, u)$ of a collaboration network. Due to limitation of the data-mining techniques to automatically extract full history of scientific employment, we assume that a scientist has one affiliation during the four year period.

The data set for our model consist of 3639 Dutch authors with 749 different institutions. However, the theory underlying our mean-field model requires that the population of each class should be large enough to be approximated by the law of large numbers. To this end, we applied an abstraction on classes (institutions) based on statistical metrics for the given distribution $D$ of computer scientists among institutions.

Since both our data set and results are focused on the system of Dutch computer scientists, we distinguish

(1) institutions in the Netherlands, and

(2) institutions in other countries.

For each of these categories we estimate a statistical threshold of the significance of the institution. This threshold depends on the dispersion of the distribution $D'$ of scientists sampled for each of the categories of institutions. If values are highly dispersed, then we set the threshold to be the average number of affiliated scientists.

To measure the statistical dispersion for the scientists' distribution $S$, we compute *a sample covariance*, which is the average distance to the mean value between any two values in the distribution $S$. To allow for some dispersion, we compare the arithmetic mean for $S$ and its sample covariance: if the sample covariance for a subset $S \in D$ is higher than the mean, then the values of the sampled $D'$ are highly dispersed.

In addition to estimation of the significance threshold, this simple test is applied in two steps:

(1) for the continental abstraction, and

(2) the country-wide abstraction.

In case 1, we sample data for all universities per continent (using the UN list of countries per continent and GPS coordinates). In the case of high dispersion in the number of scientists in institutions in one continent, we proceed to test the dispersion of the number of scientists affiliated with institutions in one country. We merge only those institutions that have a number of scientists below the mean of the entire distribution $D$. The histogram in Fig. 5.1 shows the number of scientists in each class, before and after the classes abstraction. The number of classes has been reduced from an initial 749 to 157, effectively reducing also the state-space size.

### 5.5.4 Other parameters abstraction

**Scientific age** The scientific age $h$ is based on the first publication date of an author according to DBLP. The earliest possible publications in DBLP date back to 1971, which inevitably leads to an increase by a factor 40 of the state-space size of our model. Considering our sampled data set with only 3639 scientists, the distribution of the population in such a state space is very sparse. Thus, we identify five main groups of scientific age, categorizing age into ten-year periods

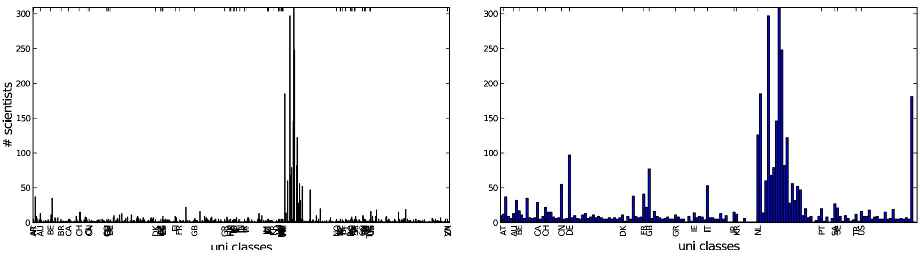



Figure 5.1: The distribution of scientists among institutions before (left) and after (right) the abstraction

as to generalize about generations of scientists: $70$, $80$, $90$, $2000$, $2010$. In general, scientific careers require substantial investments to establish tenure. These positional differences, whether it being established tenure, or a starting PhD, all influence the manner in which scientists undertake collaboration (Bozeman and Corley, 2004; deB Beaver and Rosen, 1979). Our abstraction granularity is fine enough to strongly indicate the scientific position of researchers, e.g., senior staff, junior staff.

**Visibility** The visibility of the scientists is measured by the annual number of conference publications. We choose only conference publications, as a potential interaction point, assuming that scientists encounter future collaborators during conferences. Without loss of generality, we limit the highest number of conference publications per year to 12 assuming it takes on average one month of preparation per publication. Those scientists that publish 12 and more papers per year we distinguish as fast publishers with a parameter value of $12$.

**Cosmopolitanism** The cosmopolitanism of the science is measured by number of coauthors, indicating how well connected a scientist is. We studied the distribution of the number of coauthors on our sampled data set. We observed that there are few publications with a large (more than 12) number of coauthors on a single paper. A high number of coauthors on a paper generally indicates a participation in a large research project. This results in an unnecessary large state-space size of the model, given the sampled authors in this sample. To tackle this, we distinguish five categories of coauthor count per paper: "non cooperative" ($0$) for the papers with one author, "regular" (1) for the papers with up to $3$ coauthors, "high" ($2$) with up to $6$ coauthors on the paper, "team" ($3$) with up to $10$ coauthors, and a "large project" ($4$) for papers with more than $10$ coauthors. Since we consider the unique coauthors of a scientist as possible network contacts within one year, we take the annual number of coauthors relative to the number of the publications per year per person.

### 5.5.5 Transitions and Distributions

There are three categories of distributions needed to derive from our data set for our mean-field model:

(1) communication $\kappa$,

(2) idle $\eta$, and

(3) collision $\phi$.

*Communication* is defined as collaboration via shared coauthorship between two scientists resulting in a conference paper. Both *idle* and *collision* states signify the decay of communication; in fact, for our application, these probability distributions are both an identity function. Moreover, in terms of the model, selection of the collaboration partner is governed by the distribution function *contact*, which specifies the collaboration network topology.

**Computing transition probabilities** We first measure from the collected data the evolution of collaboration between scientists (nodes) for each year 2006–2010. That is, we compute the state vector $\delta(t)$, entries of which are the fractions of nodes in every possible state of the system at time $t$. This state vector $\delta(t)$ is used in the initial configuration for the model: we sum up all fraction of nodes with scientific age $h$ from class $u$, $\delta_{(p,c,h,u)}(t)$ for all possible $p$ and $c$ and set the result as $\delta_{(0,0,h,u)}(0)$ at the beginning of each year $t$. In the model, we split the time frame onto a week $\tau$, for finer granularity, with 52 weeks in each year.

Consider states $A = (p_a, c_a, h_a, u_a)$ and $B = (p_b, c_b, h_b, u_b)$. For each pair of classes $u_a$ and $u_b$, we compute the probability *contact*$(u_a, u_b)$ that a node from $u_a$ contacts any node in $u_b$ in year $t$ as follows. Each paper $i$ with $c_i$-authors by a node from $u_a$ and a node from $u_b$ gives the probability $P_i(c_i, u_a, u_b) = \frac{1}{m(u_a) \cdot c_i}$ that the node from class $u_a$ contacts a node from $u_b$. Here, $m(u_a)$ is the number of nodes in class $u_a$. Since we have to take into account that papers jointly written by nodes from $u_a$ and $u_b$ may have other coauthors, divisor $c$ distributes the share of contribution to each coauthor. Then, *contact*$(u_a, u_b)(t)$ is obtained as follows: *contact*$(u_a, u_b)(t) = \sum_{i(u_a)} \sum_{i(u_b)} P_i(c_i, u_a, u_b)$, where $i(u_a)$ and $i(u_b)$ means "for each author of paper $i$ from class $u_a$" ($u_b$, respectively).

The computation of the collaboration distribution $\kappa_{(A,B)}(t)$ is as follows. For each paper penned by authors in states $A$ and $B$ (within a one-year time frame), we observe all possible state transitions (i.e. before and after collaboration). The result is an expression of the form:

$$\kappa_{(A,B)}(t) = \{(p_1, (A, B), (A_1, B_1)), \ldots (p_n, (A, B), (A_n, B_n))\}$$

where $p_i$ is the probability that the nodes in state $A$ at time $t$ make a transition to state $A_i$ at time $t+1$ (and, those in state $B$ move to state $B_i$, respectively). All these distributions are normalized to a weekly timescale.

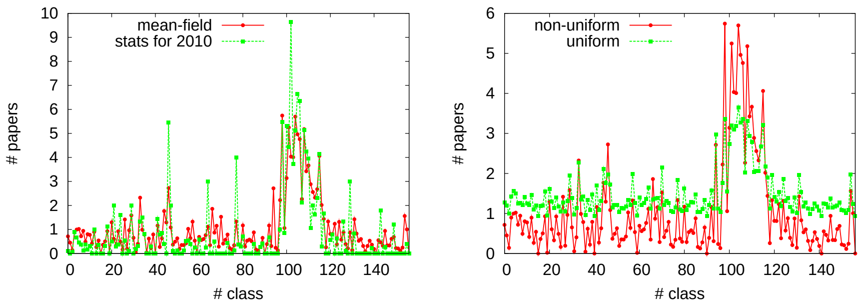
**Estimating distributions** These rates may vary from year to year thereby requiring an average to be determined for every of these distributions to ensure accuracy in the model. To that end, we obtained probabilities, as described earlier, for the years 2006–2008, and use an HMM approach to sample the underlying distribution. Our goal is to approximate the set of pairs that have positive probability of collaborating. Our mean-field model takes these sampled distributions as its input.

## 5.6 Results

The mean-field model allows us to predict average behavior. The analytical results to the statistical results for the years 2009 and 2010 are compared to the ones produced by the mean-field model. Institutions are labeled and sorted in lexicographical order; this list is enumerated and corresponds to the number on the $x$-axis (similar to Fig. 5.1). Classes 98–116 correspond to Dutch institutions. As we can see from Fig. 5.2a the mean-field results for the larger institutions corresponds with the statistics from the data set for 2010. Our data set does not list all papers of the coauthors of coauthors, but we divide by all people in the class; so statistics produced are lower than actual.

**Institutional factor** The results produced by the alternative mean-field model with uniform distribution *contact* for collaborations between different institutions show that the sample distribution is non uniform. This *contact* distribution produces the equal probability of collaboration between any two scientists in the whole network, irrespective their affiliations, and thus forms a baseline for comparison to see whether affiliations are statistically significant. The comparison in shown in Fig. 5.2b. As we can see, the uniform *contact* distribution predicts higher output for foreign institutions but lower for Dutch institutions, since the output is then uniformly "redistributed".

**Impact of scientific age** Fig. 5.3a shows the average number of papers for different scientific age. The results from only Dutch institutions were averaged. The mean-model model shows that a principle of preferential attachment (Albert and Barabási, 2002) is occurring in the network based on age, with higher tenured scientists acquiring more collaborators and papers. The average output




(a) Comparison with mean-field. (b) Comparison of different *contact*.

Figure 5.2: Average output for different classes.

| Sci. age | avg # pubs. |
|---|---|
| 2010s | 1.8 |
| 2000s | 1.61 |
| 1990s | 1.76 |
| 1980s | 1.95 |
| 1970s | 2.3 |

(a) Average output for different scientific age.

| $u_a \leftrightarrow u_b$, $u_b \leftrightarrow u_c$ | avg. $u_a \leftrightarrow u_c$ |
|---|---|
| $>=$ 0.0 | 1.0 |
| $>=$ 0.2 | 1.13 |
| $>=$ 0.4 | 1.15 |
| $>=$ 0.6 | 1.20 |
| $>=$ 0.8 | 1.27 |
| $>=$ 1.0 | 1.32 |

(b) Triad relations.

Figure 5.3: Results for the age impact and triad relations for Dutch institutions.

per scientific age per institution, was also computed, which displayed differing tendencies in collaboration patterns.

**Link prediction** In accessing the manner in which links are made through transitivity: if class $A$ has a paper in common with $B$, and class $B$ with $C$, then $A$ has stronger connectivity with $C$. Within this system we consider the institution parameter, allowing us to reflect on the initial hypothesis – an institution plays a role in the collaborative patterns of scientists. The connectivity factor based on the distribution *contact*, which in turn, depends on the probability $P_i(c_i, u_a, u_b)$, the number of coauthors from a certain institution implicitly contributes to strength of the connectivity between institutions. Fig. 5.3b shows the generalized triad relations of Dutch institutions; considering a scientific age in *contact*.

## 5.7 Discussion and Conclusion

In investigating the system of Dutch computer scientists' collaborative behavior through the mean-field model we observed systematic networking behavior associated with a number of social parameters, which aid in describing the networking dynamics of scientists. The past collaborative partners of one's institution plays a key role in how future collaborations unfold. With every conference proceeding with another institution the chance of collaborating with the institution increases. Age also matters; the age of the scientists plays a role in the visibility of a scientist (number of publications) within the system. The cosmopolitanism of the scientists (number of co-authors) also contributes to the likelihood of future interaction. Consequently the mean-field model allows us to describe the Dutch CS system of conference paper collaboration to be governed by a number of social variables, where ties can be predicted given previous relationships among common institutions, reinforcing clustering tendencies in these networks.

In this first application of the mean-field model in predicting both social and network parameters for large social networks, we also recognize a number of shortcomings. The first is the sensitivity of the data-aware approach and thus the empirically informed aggregations of nodes into clusters from such an approach. Future work should aim to consider additional social parameters, such as performance, gender, discipline, length of time known in understanding the system. To improve the precise description of states the notion of idle and collisions in the model should be improved for social networks. Additionally, we acknowledge that this explorative study of the mean-field model did not address both the potential for shift classes reflecting the fluidity of actual organization constraints in social life, as well as model checking. These limitations are related to the current state of computing techniques, in first data-mining techniques which does not currently allow us to collect such refined information on social beings, and secondly the lack of methods to appropriate accurate model checking.

The incorporation of the modeling knowledge with population specific dynamics we are able to identify the conditions under which links emerge given a set of both network and social parameters through the mean-field model. This allows us to provide informed predictions to comment on the mechanism(s) under which specific patterns of behavior emerge in large social networks. Mean-field models provide a meta-scopic method, which overcomes limitations of the network only and social parameter models. Meta-scopic models of this sort allow us to incorporate both the micro (considered in evolving graph models) and the mega networking processes to infer links through a data-aware approach. Additionally, it provides an empirical terrain at which to explore the effects of both network and social parameters on large social networks.

# 6 Conclusion

This dissertation research investigated the effects of exogenous contextual factors on the success and emergence of social network structures. Studies of network consequences suggest that context has an effect on the success of specific network structures, and that our knowledge of these structures is incomplete without a look at antecedents to explain the emergence of structures. Thus, this research specifically investigated the role of a set of contextual factors as a determinant in explaining differences in network structures, as well as the success of the networks employed by individual researchers involved in scientific collaboration.

In undertaking this research, I reviewed theory on networks which outlines three generative roots to network emergence– endogenous network factors, individual factors, and exogenous contextual factors. I explained how theories of the role of these roots– the *modi operandi* of social network dynamics – contain a blind spot in the consideration of exogenous contextual factors. The lack of a conceptual framework that explains the interplay of context and individual actions in understanding network emergence leads to the lack of a methodological approach for investigating possible effects. I proposed the integration of structuration and network theory to delineate the role of context on network structures and outlined a framework from the concept of *duality*. *Duality* is the interplay between rules and resources that explain how structures employed by individuals realize actions. Using this theory, I conceptualized the role of the three roots. I considered rules to be key contextual factors, as identified through policies in nine different academic departments as an interplay with resources . These endogenous network factors and individual factors were measured from previous network positions and attributes of the individual researcher. This allowed me to explore how context influences the emergence and success of specific scientific collaboration network structures.

In pursuing this research, a mixed-method model research design was employed which allowed me to delineate possible contextual effects as a moderator and develop a better understanding of the emergence and success of individual researchers scientific collaboration networks. I implemented three separate models, thereby investigating effects at three levels of analysis - individual, organizational unit, and the field – in three empirical chapters. The key findings that emerged from this research are as follows:

(1) The policies of a professional tenure system and a publication target list within a department enhance the success of cohesive scientific collaboration networks of researchers (See Chapter 3).

(2) The existence of a tenure system influences the generative roots that explain the emergence of these scientific collaboration network structures of a department (Chapter 4).

(3) Tendencies to collaborate can be explained by a common university affiliation (Chapter 5).

I elaborate on these key findings and how they contribute to current knowledge in Section 6.1 below. These findings have a number of implications for our understanding of scientific collaboration, which I discuss in the section on Practical Implications (See Section 6.3). I also discuss a number of limitations and suggestions for future work. I bring together these findings in the final section of this chapter (see Section 6.5) to explain how this research contributes to the specification of the role of exogenous contextual factors as a determinant in explaining network structures.

## 6.1 Findings

The empirical chapters of this research dissertation looked at three different levels of aggregation of scientific collaboration networks in order to investigate the effect of context as a determinant on observed scientific collaboration structures. Context was identified through the identification of a number of policies that attempted to steer publication within nine Dutch Computer Science departments. Scientific collaboration structures were measured through publication data, inferring collaboration through co-authorship. Information on individual researcher attributes was garnered from both Web data and survey data depending on the model used in the empirical chapter. The success of realized publications of the individual researchers was measured through raw delayed citation scores. I discuss here the key findings from the three empirical chapters.

Findings in Chapter 3 showed that the policies of a professional tenure system and a publication target list within a department enhance the success of cohesive ego network's scientific collaboration. Specifically, these two policies presented a set of rules that stipulated core targets for researchers' publications and how their publications would be evaluated. Researchers in departments with both a tenure policy and a publication list achieved enhanced success in implementing scientific collaboration networks. Researchers in departments with other combinations of policies, or a lack of the two policies, saw enhanced success in less constrained networks. The highest rates of success were found in constrained networks where co-authors shared collaboration, compared to less constrained networks where a researcher is situated between co-authors.

To investigate how these structures are realized within the department, I explored how contextual factors relate to the specific emergence of generative roots as an explanation for different network structures in Chapter 4. I found, through the implementation of a recasting of the traditional use of SIENA models, that indeed different generative mechanisms are at play in the nine departmental units. Findings showed that the existence of a tenure system influences the generative roots and explains the emergence of scientific collaboration network structures of a department. Researchers in all of the academic departments have a tendency to initiate and maintain collaboration ties that are transitive. There is a tendency to initiate and maintain collaboration ties with popular authors, which leads to the so-called Matthew effect (Merton, 1968) of preferential attachment (Albert and Barabási, 2002), where researchers with a high numbers of collaborations increase further collaboration exponentially. Tendencies to connect with actors with high *betweenness* – the holding of a brokering position with greater potential to control communication and information within the network (Burt, 2005) – vary in both significance and direction of effect between the departments. Within all academic departments, we find that researchers have distinct tendencies to connect with co-authors, and the strengths and existence of these effects vary between the departments. Departmental units with a tenure system, in particular, have tendencies toward transitive ties, as well as a negative likelihood to collaborate with researchers who occupy broker positions, thus confirming findings in Chapter 3, where cohesive networks are more successful in these contexts. The opposite effects are found in departmental units with no tenure system, suggesting that the emergence of different structures is contingent on the policies in the context in which the individual is embedded.

In considering the aggregated network behaviors of individual researchers to identify whether exogenous contextual factors play a role in guiding tendencies to collaborate in a field, I tested a mean field model in Chapter 5. The mean field model allowed me to identify the conditions under which networks emerge through the averaging of network behaviors. I considered a number of variables to explain the emergence of collaboration. Findings showed that tendencies to collaborate can be partly explained by a common affiliation. The age of the scientist plays a role in the visibility of a scientist (i.e., number of publications), as well as the number of co-authors, which also contributes to the likelihood of future interaction. The affiliation of past co-authors plays a key role in how future collaborations unfold; the addition of every publication increased the chance of future collaboration with another unit.

## 6.2 Contributions to theory

Findings from all three empirical chapters showed that context plays a role in explaining network dynamics. When considering exogenous contextual factors as a moderation variable in explaining network structures, rather than a boundary condition or control variable, we are provided with evidence to suggest that it enhances the likelihood of the emergence of specific generative roots and network structures.

These findings contribute to the ongoing debate on the efficiency of constraint as viewed as a continuum from structural holes to cohesive networks (Burt, 2005; Coleman, 1988; Granovetter, 1985). Research has suggested the context plays a role in network performance. Given that we can assume that the tasks that researchers are pursuing are complex and specialized in nature, this research confirms findings that high constraint in networks is effective for completing complex knowledge tasks (Cummings and Cross, 2003; Hansen, 1999; Reagans and McEvily, 2003), with the condition that the organizational unit has policies that seek to steer outputs. This is in contrast to Burts (2000) proposition that constraint is an advantage in larger groups with low uncertainty. Thus, within knowledge-intensive activities such as scientific collaboration, policies that attempt to avoid uncertainty in outputs alter the way in which researchers realize publications via potential collaborators, making cohesive networks more successful. Future research should work to specify these conditions and test a contingency model where particular sets of factors relate to network success, as well as to further identify sets of factors that influence the success of different network structures.

Given that context can be seen as a dependency for social mechanisms, a number of implications can be garnered for current theories on network dynamics. In this research I only considered one type of network; thus it is fair to confirm that the strength of context as a determinant remains conditional for the network relation under investigation, as proposed by Shumate and Contractor (2013). The determinant role of context puts to question in particular Whitbred and authors (2011) work, which suggests a less significant role for external contextual factors.

In this study, I identified contextual factors through the existence of policies and clearly defined a context as an organization or a community with different sets of rules. Additionally, this process included the formalization of rules pending their stage of development, thus the resulting effects would be different; suggesting different *modi operandi* of networks per context. If context remains to be considered as a boundary and not as a determinant to the emergence of specific locally observed mechanisms, then indeed we may be over-attributing the role of other roots in explaining dynamics. Overall this brings to light the necessary further theoretical exploration of exogenous contextual factors in constraining and/or facilitating the emergence of other roots and, ultimately, other structures. Future studies on network dynamics need to consider multiple contexts in exploring network dynamics. These findings further put to question Ahuja, Soda and Zaheer's (2012) proposition regarding individual microactions as leading dynamics.

In summary, the context within which researchers develop and maintain their scientific collaboration networks influences their success, as well as the behaviors employed to realize these networks and the tendency to collaborate. This evidence has implications not only for how we understand dynamics but also for identifying the conditions within which individuals can employ their networks in a specific manner towards a number of possible outcomes.I discuss these here further as practical implications.

## 6.3 Practical Implications[p]

This research focused largely on theoretical and methodological advances necessary to investigate the determinant role of context on network structures. Findings also have a number of practical implications for the case of Dutch Computer Science and science policy in general. These implications are seen as recommendations drawn from the findings of three empirical chapters.

The findings of this research suggest that context directs the networking behaviors of individuals. Specifically, policies implemented by Dutch Computer Science departments partly explain the emergence of the specific network structures of individual researchers' scientific collaborations, as well as the success of these collaborations. I focus here on context as defined through the existence of a professional tenure system and a target publication list. Tenure systems facilitate a set of guidelines for evaluation and promotion. These are contractual obligations that have consequences; thus it is in the researcher's interest to most effectively produce outputs to maintain or achieve positive evaluations and increase formal status. A publication list creates a heightened awareness of a target for knowledge outputs. Such target lists are used to steer day-to-day conditions stipulated by the formal organization that facilitates employees' work. These are often more specific and detailed lists of goals, or evaluation mechanisms, that attempt to steer specific behaviors and/or outcomes. The combination of the two provides a set of rules that defines accepted behavior; in this case, the conditions under which publications are evaluated and incorporated into professional and promotion decisions are the result.

Findings from the three empirical chapters that make up this research showed similar scientific collaboration networking behavior by researchers with common affiliations (see Chapter 5). Context, and the existence of a tenure system in particular, promotes different behaviors that explain the emergence of collaboration networks among researchers (see Chapter 4). Additionally, the success of the scientific collaboration networks implemented by researchers differed, depend-

[p]This section is part of a working paper in cooperation with the Rathenau Institute on Dutch Computer Science, in combination with Chapter 2.

ing upon the existence of a tenure system and a publication list (see Chapter 3). Researchers anticipate enhanced success with cohesive scientific collaboration networks that are housed in departments where both specifications for professional tenure and a list targeting publication outputs exist. In these contexts, scientific collaboration networks, where more co-authors shared relations, resulted in greater publication success. In the case of a department that has less specified guidelines, whether because of the lack of specific policies or the existence of just a single one, a less constrained network is more successful. These are networks where the individual researcher sits between other co-authors or co-author groups, serving as a type of broker between co-authors. These policies alter the way in which researchers realize publications via potential collaborators. A strategic position in one context requires a different network than in another. Researchers are facilitated and/or constrained by such specific policies about outputs and thus strategize accordingly. Thus, social selection mechanisms are dependent on the context.

### 6.3.1 Policy Recommendations

In addressing policy recommendations from this research, I focus here primarily on the findings that suggest that departments with both a tenure system and a publication list enhance the success of cohesive scientific collaboration networks, from Chapter 3. These policies outline a set of specifications that guide the behavior that researchers evoke to achieve publication through collaboration. This leads to distinct networking strategies where researchers in departments with specifications can focus on their local positions (i.e., within the department) and invest in cohesive teams, whereas those without a list or tenure track system, or with uncertain criteria, position themselves more globally. Currently thinking on publication success has been attributed, among other things, with top ranked universities, where top ranked departments attract better researchers which work together to explain the increase in productivity (Allison and Long, 1990). Rather these results question the role of departmental policies, not just the organization as a site of production, as an explanatory factor for scientific collaboration success. Thus, given the stipulations via policies researchers employ specific behaviors to meet professional expectations which influences the success of their collaboration decisions. These systems act as herding system where those that do not behave this way are effectively removed or leave.

The success of cohesive networks in departments with publication steering policies are vulnerable to the increasing scarcity and uncertainty of tenured research positions. This is a threat to the overall quality of both a discipline and science system. The effect of these policies come into question when the incentives of promotion from successful publications become limited or even nearly impossible. Policymakers should adhere to the importance of balancing relatively cheap, junior research positions through funding mechanisms, departmental reorganizations and the like where tenure becomes only achievable for the very few; with funding for longer term research projects, and permanent chairs to sustain a healthy system.

In addition, in regards to Dutch Computer Science as a field. Despite that conference proceedings comprise the largest share of publications and are also among the highest cited publications within computer science, they still remain often absent from commonly used indexes and formal reviews and evaluations by faculties and universities. This distinct communication practice is often misinterpreted as a way to publish low quality or findings of less theoretical or practical value; when in reality these proceedings are significant products that serve both as markers of the patterns of collaboration that researchers undertake to produce knowledge, as well as the attributed quality of knowledge. Evaluators should take note of the impact of evaluated knowledge outputs of multiple forms and the networks that researchers employ in developing scientific knowledge when considering candidates for positions, funding and reviewing quality.

From these findings, I elaborate here on a set of recommendations for steering both contextual factors and scientific collaboration networks in achieving publication success from the perspective of researchers, institutions and or managers of research within organizational units, as well as funders.

### 6.3.2 Researchers

This research has shown that there is interplay between the rules stipulated by academic units for steering outputs through evaluation mechanisms and resources of the individual to realize collaboration through publication. The policies of a professional tenure system and a publication target list within a department enhance the success of cohesive scientific collaboration networks of researchers. For example, if a researcher works in a department which has both a publication list and a tenure system, a highly constrained collaboration network would be optimal. This means that the researchers' collaborators are also relatively well connected among each other. If, on the other hand, the department has neither publication list nor tenure system, our findings suggest that the researcher might better build a network saturated with structural holes, e.g., low constraint. This means that the researchers collaborators are not particularly well connected among each other. These differing contexts yield different sets of constraints that influence the success of collaboration network structures.

Researchers themselves should work to be aware of both the structures of their collaboration networks and the policies that aim to shape outcomes through evaluation. Individual scientists can try to build their collaboration networks in accordance with the departmental policies and their own goals. In order to successfully navigate a tenure system, it is important to recognize evaluation criteria. The same holds true for departments that lack a formalized tenure system or of a publication list that aims to target publication outputs, although less explicit.

As a suggestion, a researcher should be aware of the why they are producing knowledge through collaboration. A first step in doing this is being aware of those you collaborate with, and also how you collaborate with others to realize publications. If you realize your publications with a core team, work to facilitate their communication and be sure to insure injections of new knowledge to maximize knowledge exchange and success through publication. If you function as a broker between different groups of coauthors, that do not collaborate with each other through publication, be aware of how you share knowledge with the different groups, and reflect on your position given different constraints of projects. Thus, when considering one's collaboration networks, scientists should contemplate not only how to stimulate interaction between co-authors, but also how to interject new knowledge or additional co-authors in order to reduce the risks of homogeneity of knowledge within their co-author group.

A number of tools exist that visualize scientific publication networks (e.g. socio-grams) and serve as effective instruments for thinking about how to collaborate on future projects. These networks should be reflected upon in discussions with other researchers, department heads, and or other academic managers to understand how these collaborations can be facilitated within the constraints of the department. Consider how you will develop and maintain relationships with need expertise and future co-authors when starting new projects to ensure and encourage communication between all authors. This advice is also pertain for those in departments without tenure systems where successful scientific collaboration network structures are more sparse and individual researchers act as brokers between other researchers in realizing publications.

### 6.3.3 Academic Institutions

These findings provide suggestive evidence about how organizational units such as departments and universities steer success within a field, not necessarily through physical or monetary resources, but through policies that promote specific behaviors. Thus the conditions of these academic units serve as a selection system for cultivating a specific end. Given these findings, the institution plays a key role in the publication success of these scientific collaboration networks.

Thus, I recommend that academic institutions should first work to stimulate interaction between researchers and encourage collaboration in sub-groups for the building of cohesive collaborations and successful publication. Institutions should also keep in mind the nature of dependencies, given researchers' subfields, as some fields have a greater propensity to interact given the research domain. Institutions, therefore, should act accordingly to insure that researchers are provided with the tools and access to other potential co-authors, in order to decrease the risks of so-called "group think". They should encourage regular contact with others outside of the researchers' core scientific collaboration networks in order to stimulate new knowledge and ideas for reinforcing the potential success of these networks. Activities such an affiliation with a research institute and access to industry should be stimulated and appreciated at the same level as outputs.

In order to enhance the success found in these cohesive networks, institutions should stipulate both a tenure system and a publication list. For those units without a tenure system, investments should be made in developing one. The criteria of these tenure systems and the content of the publication target lists are unique to the context, and they serve as a guide for steering behaviors to realize collaboration through publication. In particular, institutions should take care when considering success within these policies. In this research I solely looked at performance of publications through citations, although there is an increasing number of ways to measures success in science. For example, journal rankings, author orders, societal impact, alt-metrics, interdisciplinarity, and so forth can be considered. Just as conditions determine success, so do the necessary behaviors and contextual factors of different innovation policies. These measures are not "one-size-fits-all" for they are dependent on the goals of the academic institution, faculty and department.

The department as a unit plays a key role in negotiating how these policies are received; thus, managers or management teams need to be proactive in shaping these policies through commissions or review boards. The development of target lists for outputs should be an ongoing discussion among groups in defining the impact of the researchers within the department. Additionally, for the discipline of computer science, it is imperative that conference proceedings hold an equal or nearly equal value in evaluation criteria, particularly if tenure policy is a faculty-wide one. The lack of consideration of proceedings for computer science, particularly in evaluation criteria, leads to inaccurate measures of the total impact that publications have in contributing to scientific knowledge. It puts computer science researchers in an unfair position as it discredits the majority of their work.

As, mentioned above, the success of these policies is dependent on the promise of tenure. Recent dramatic decreases in academic funding are no doubt a threat to the capacity of research in general. The promise of tenure provides motivation to invoke different collaboration strategies for collaborative publishing relationships, without this the incentives loose effect. Academic institutions should work to insure professional mobility of researchers to sustain quality research.

### 6.3.4 Research Funders

Increasingly, research grants demand cooperation (Defazio et al., 2009). This has been implemented under the guise that collaboration or, rather, interaction – is positive for the stimulation of knowledge, and, certainly, that is the case. However, as this research showed, the local conditions under which collaboration is realized have an effect on its likely success. This makes for a challenging landscape for researchers, as there is a danger that criteria for funding mechanisms abruptly shifts in juxtaposition to the conditions of the local context; where perhaps cohesive networks are most successful at the cost of funding demands to cooperation with international partners. This may be in direct opposition to the strategies necessary for successful publication due to barriers of infrequent physical interaction and electronic communication which then run the risk of fostering distinctly different networks. For example, this research suggests that the likely infrequent physical interaction, with possible conflicting local mechanisms between shared grant holders may lead to less successful sparse, low constrained scientific collaboration networks. Thus, the decision to fund large cooperation projects has implications for both the scientific collaboration networks built by researchers and the over success of the science system. In particular, funders should evaluate whether these international, or multi-party, cooperation requirements truly reduce the exchange barriers through the building of cohesive groups of co-authors; as well as adequately fund the opportunity for exchange through travel to stimulate network building.

## 6.4 Discussion

This research is not without limitations. I reflect here on these limitations in general; more specific limitations to data collection or methods are discussed in detail in each of the empirical chapters. These factors include the limitations of generalization of the case to other fields in science, data collection validity and reliability issues, as well as a discussion on the use of methods.

### 6.4.1 Dutch Computer Science as a Case

The case of science provided a valid and a reliable setting to explore social network dynamics, as I was able to validly identify boundary conditions and potential individual characteristics given the large body of literature on science studies. This case and the findings of the research are generalizable for many other fields of science. However, they are applicable largely to fields that have high turnaround rates for publications, and thus produce knowledge at a high rate, given the nature and importance of innovations for daily life. For example, we may think of medical science as a comparable case, but not necessarily the humanities, given different research techniques and knowledge turnaround. Findings are not necessarily generalizable for fields in which single-author and small author groups are the norm, therefore.

An independent tertiary review of the field (Nationale Informaticakamer, 2010) of Dutch Computer Science provided a valid selection of cases to distinguish nine academic departmental units conducting academic research. This provided clear boundaries for identifying rules within contexts, as well as conditions for identifying a set of individual researchers. Within each of these departments I interviewed both experts and the head of each department to identify the conditions under which researchers were working. I attempted to overcome possible subjectivity through the use of semi-structured interviews and context analysis of the interviews in order to identify overlapping issues in describing the field.

In increasing validity, future work should seek to identify factors within the context via tertiary data, or possibly by studying online communities in which these conditions may be explicitly set, given a context of rules (Williams et al., 2011).

### 6.4.2 Data

In addition, I recognize a limitation in regards to the recall of interviewees and respondents, as discussed in Chapter 3. Researchers were asked to reflect on scientific collaborations emerging from 2006–2012, as well as conditions of the context from 2005–2010. Considering the data collection period, this related to a period of approximately five years, and was asked about one to two years later (the survey occurred after the interviews). This decision was made to capture the scientific collaborations that likely came out of this period, allowing for the capture of a lag in final publication and citations. For example, if I were to start working on a paper today in 2014, it may not be realized for months or published for years. Given that the largest portion of data asked in interviews was largely static (e.g., gender, year of tenure, promotion), and the emergence of formal policies takes time to formalize and be initiated, a number of recall issues can be mitigated.

The majority of data used in this project related specifically to the dependent variable of networks. The quality of data and the network data set, in particular, are the strengths of this research. The network data were queried from public sources, and thus do not present an issue of recall (Moody, 2004). This publication data was also used in computing variables. The reliability of historical publication data is related to the reliability and validity of the data store used. The publication data used in this research was queried from DBLP (Database Systems and Logic Programming, 2013). The use of this data store has advantages, as disambiguation issues are reduced, given that DBLP provides a unique ID per individual name and researchers have a tendency to self-monitor their DBLP pages.

This leads to increased data quality. Given this accuracy, I was able to assume complete datasets of all scientific collaborations. I did not need to consider simulating missing data, which provides a certainty in affirming the network structures and, consequently, the results. This increased the validity of variables and reduced the dependency on individual researchers, which is traditionally done in social network studies. Future work should investigate the use and biases of Web sources in querying assumed complete datasets of individuals active offline (see work completed during this dissertation project for the directions and implications of this type (Birkholz et al., 2013)).

With respect to the consideration of publication data, I assumed affiliation, inferring relationships between researchers via a shared event. Given the reliability of the database where the publication data was queried, I did not employ a second control in confirming a relation. This assumption held largely true as the majority of researchers maintained relatively manageable collaboration relations (see Descriptive Results in Chapter 3), with a few exceptions of researchers with co-author lists in the hundreds. Additionally, the data was dichotomized in Chapters 4 and 5, meaning I did not consider weighted or valued networks. Dichotomization is a common practice done in implementing SIENA (Snijders et al., 2010; Whitbred et al., 2011). Because I was interested in the structure of the network that emerged, not the strength of these ties, such an approach was appropriate. Future work should seek to expand this toolbox by considering weighted networks in social network dynamics, following work by Opsahl and Panzarasa (2009).

### 6.4.3 Methods

Given the blind spot acknowledged in this research, a mixed-methods research design was implemented in this dissertation to explore a number of possible influences in understanding the effect of context on network structures. This approach entailed methods that allowed the investigation of effects related to both antecedents and consequences, as well as static and dynamic models that considered context as a moderating variable given the current state of methodological tools available as described in detail in Chapter 4.

Chapter 3 implemented a moderation model of aggregated static ego networks to investigate the effect of a number of policies of academic departments on the success of individual collaboration networks. Findings confirmed that contextual effects have a consequence for network success, providing evidence to question the antecedents. In investigating antecedents, Chapter 4 took a dynamic approach considering network panel data of departmental units in influencing the emergence of generative mechanisms to explain network structures comparing nine SIENA models. The assumptions of the SIENA model provided a strong theoretical model for the investigation of network dynamics and to identify generative roots, although it also presented a number of limitations (e.g., computational limitations for network size) that do not allow for the consideration of an exogenous context as a predictor in explaining different network structures. In an attempt to overcome these limitations and still take advantage of the strengths of the model, I presented a recasting that qualitatively compared SIENA models in order to identify patterns related to the context. A number of limitations remained. The greatest limitation for implementing such models in exploring the effect of context remains the study of closed networks  a classroom, a school, an organization. Thus, although the recasting of SIENA as proposed here would allow us to test if context plays a role in the emergence of relationships, it fails to capture how context potentially influences the emergence of different generative mechanisms that could allow an explanation of differences in structures. Additionally, it fails to consider inter-network interactions, when in reality these networks may overlap. Considering networks with known overlaps is a more valid approach as it undertakes a meta approach and considers all behaviors of these individuals, as constrained by rules of the context. It also allows the consideration of effect that have the potential to interact within a field of others who are embedded in different contexts. I experiment with the implementation of this model in Chapter 5. I discuss the limitations and future work of this model specification in the following section.

As proven in this research, the use of the mixed-method model was an appropriate, necessary and valid choice to explore various possible effects of context. Such a toolbox allowed me to provide a broad set of evidence at which to understand the potential effects of context. This study also contributes to growing knowledge on the development of mixed methods for studies investigating both context as an antecedent as a determinant to the emergence and success of social network structures in studying large social networks. Thus, it suggests that effects need to be investigated at multiple levels  the ego, the bounded, relatively small, organizational context most often explored in social science studies of networks and a more global field approach. This allows the exploration of the complete reach of effects in explaining network antecedents. Future work on network dynamics, and particularly in further exploration of context as a determinant in explaining network structures is in need of such studies to outline the reach of effects at which to be able to better theorize the conditions under which particular network structures both emerge and are successful.

**The mean field** A first step in overcoming a number of methodological limitations involved in the consideration of exogenous factors was the implementation of the mean field model for investigating social network dynamics, as described in Chapter 5. This was the first use of a mean field model in explaining collaboration tendencies. Through the implementation of the mean field model I was able to consider a larger social network (or organizational boundary) in the identification of the tendencies of scientific collaboration. This allowed me to overcome limitations, as presented in Chapter 4, through the averaging of behaviors to explain network emergence. Further, I considered the context as an explanatory variable in identifying dynamics, instead of a boundary condition, as in similar models.

The mean field model was historically used to define complex dynamics for large systems. The use of such a model that overcomes computational limitations through the averaging of behaviors of the system allows for the identification of factors that play a significant role in network dynamics. In implementing this aggregation, some information is effectively lost (e.g., the SIENA models in which I could not identify the precise generative mechanisms within these contexts). The use of this model thus highlights explanations for overarching dynamics, complementing the understanding of the network dynamics identified in more common approaches. In identifying these conditions, and the potential shifts in their level of influence over time, we have a more accurate depiction of dynamics than the use of largely static methods common in the social sciences.

The strength of the mean field model is that it implements similar mathematical assumptions as SIENA through the use of the Markov chain model to simulate changes in the network. However, it also allows us to acquire a more bird's-eye view of the key factors that drive dynamics. Additionally, it captures the most common conditions under which the network emerges, instead of the most likely ones, as occurs in other network models. This averaging allows one to exclude extremes or factors that only play a role to small groups of individuals, while still considering a large number of possible factors in studying large social networks. Thus, I purport that mean field models provide a valid manner to identify dynamics, as well as an additional way to specify models for hypothesizing about mechanisms occurring at lower levels. This can explain the emergence of network structures and suggests the use of simplified models for the garnering of similar results when seeking to identify key overarching factors in network dynamics.

Although, in invoking the mean field model, a number of theoretical assumptions emerge that need to be considered for future implementation of the model in understanding network dynamics. Given that the mean field model was implemented in this research as a methodological step to explore context given its similarity to the SIENA model, I do not aim to outline the theory of the field in detail here, as that is not the purpose of this research. Rather, I acknowledge the need for future work in conceptualization of the field in implementing mean field models. The implementation of such a model for social networks has a number of implications given the lack of conceptualization of context, as explained in detail in the Introduction Chapter of this dissertation. This is the theory of the field.

The concept of the field and as it is implemented in the current model has its origin in the natural sciences, where the effect of field influences the way objects in a particular domain interact with one another. The field as a variable is exogenous to the individual and influences the social practice that is undertaken by the actor. The field itself is organized, and differential (Koffka, 1935, p. 117) particles differ in the degree and direction of charge. It is seen as an underlying organizational structure which does not necessarily influence all entities at the same rate. In implementing such thinking, the explanation of how the state of some elements change need not apply to the changes in state of other elements, suggesting that the field is not causal. Rather, the interplay of an existing factor in the field, in combination with other factors, explains the emergence of a behavior. Put more simply, we might think of such models as being able to identify the *states of elements* (or combinations of elements). For example, given the origin of the theory, I offer here an example that conceptualizes these different effects of the field: water ($H_2O$). Given the external conditions of temperature, water changes from ice to liquid to vapors. This exogenous factor of temperature limits the nature of an element's behavior. Thus, in specifying how and when these shifts in behavior occur, we can identify the conditions under which the state of an element changes. (Martin, 2003)

Field theory in social sciences is not new; Bourdieu (1998) wrote on the concept, as understood within the social sciences. More recently, Martin (2003) expanded on the need for such a theory in explaining differences in behavior, given the acknowledged lack of conceptual development on fields in social science.

The field in organizational theory encompasses a relational space (Bourdieu, 1998; Wooten and Hoffman, 2008)), where entities interact and operate given a shared meaning (Scott, 1995), and/or a shared or similar task (DiMaggio and Powell, 1983). Processes are explained as being guided by institutionalization (Scott, 1995) or engrained rules (Zucker, 1977), which entities (e.g. actors, organizations) take into account when making decisions. Agency is constrained given a set of factors (Covaleski and Dirsmith, 1988; DiMaggio, 1988). Thus, change can be explained where entities of a particular set of attributes or parameters are susceptible to different field effects; they (potentially) shift given a set of effects (Lawrence, 1999; Oliver, 1991). Thus in conceptualizing the field, we need to be able to stipulate first the boundaries, and then a complicated set of causal relationship arguments that define the possible trajectories given a set of contextual factors (which do not always relate to similar outcomes). This type of model building requires specification of contingencies at multiple levels; for example, given the population, this may include: the field, the organizational unit, the physical neighborhood, peers and individual factors.

The current under-specification of possible external effects in social network theory – and in empirical studies in particular – is a limitation to exploring how different conditions serve as antecedents and effect outcomes (Battilana and Casciaro, 2012; Greenwood et al., 2011) at which to greater specify relationships. This research is a first step in that call and provides insight into local effects that are context determinant. Future work should, however, explore further experiments with the mean field model to explore possible effects in developing a stronger theoretical framework for understand network antecedents – and consequences in particular.

## 6.5 Conclusion

In conclusion, this research investigated an acknowledged blind spot in current network studies for the understanding of the effect of exogenous contextual factors on the emergence and success of social network structures. Through the implementation of a theoretical framework using structuration, I was able to explore a potential interplay of exogenous contextual factors, individual characteristics and endogenous network factors to explain differences in network structures. Findings showed that, regardless of the level of the network aggregation, contextual factors explained the tendencies to collaborate and the success of network structures, as well as the generative roots that explain the emergence of scientific collaboration network structures.

Consequently, I purport that that exogenous contextual factors indeed serve as a determinant in understanding social network structures, as they serve a dependency role in understanding the invocation of social mechanisms by individuals. Thus, networks have different *modi operandi*, which can be explained through context; where local mechanisms, such as proximity, serves to organize individuals through the facilitation of behaviors that relate to realized network structures in taking action. It serves as a precursor for describing the effects of local mechanisms.

This sheds new light on previous network research, particularly for the large body of research that has solely selected networks as a boundary condition for investigating networking mechanisms. This is just the first evidence to suggest context as a determinant and, as described in the discussion, deserves further theoretical exploration for confirming these effects. However, if it holds true, it begs a question about the validity of a number of findings. Particularly it suggests if there is are different modi operandi of social networks which are context dependent that our understanding of other mechanisms needs to be reflected upon to classify the conditions under which specific mechanisms occur. Thus, in understanding social network structures, context truly matters.

# Summaries

## English Summary

Social networks matter. The characteristics of social networks have consequences for outcomes (Burt, 2005; Coleman, 1988; Fleming et al., 2007; Granovetter, 1985). The precise mechanisms leading to network structures are less understood. In order to study the dynamics of social networks, the mechanisms through which networks emerge are examined, bringing us to a better understanding of how this process works. In this dissertation a number of generative mechanisms are identified as antecedents that lead to network structures. These generative mechanisms are said to have three roots:

(1) network-only factors;

(2) individual factors; and

(3) contextual factors.

These roots describe the origin of these networks mechanisms. The roles these generative roots play in the generation of network structures remain debated, with theories suggesting different modus operandi (Ahuja et al., 2012; Liben-Nowell and Kleinberg, 2007; Shumate and Contractor, 2013; Whitbred et al., 2011). I argue context is overlooked, despite increasing evidence that the success of different network structures is contingent on different organizational processes (Battilana and Casciaro, 2012; Burt, 2000). In this research, I advance knowledge in the *modi operandi* of networks by specifying the role of contextual factors to explain the diverse network structures observed. To explore the role of context on social network dynamics I investigated the network structures from three perspectives:

(1) the success of individual network structures (Chapter 3);

(2) the patterns of generative mechanisms within networks (Chapter 4); and

(3) the tendency of interaction (Chapter 5).

In reviewing theory, I explained there is no existing framework that considers the interplay of contextual factors and individual factors in explaining network structures. Thus, I applied a theoretical framework using structuration theory for the conceptualization of current knowledge on the role of generative roots in the emergence of networks. I proposed using structuration to delineate the role of context on network structures and outlined a theoretical framework from the concept of duality. Duality is the interplay between rules and resources that explain the structures employed by individuals to realize an action (Giddens, 1984). I considered rules as contextual factors identified through policies as interplay with other a set of resources endogenous network factors and individual characteristics measured from previous network positions and attributes of the individual researcher to describe these differences in network structures.

In pursuing this exploration into the exogenous effect on dynamics, several methodological extensions were necessary. A mixed-methods network approach is taken to evaluate the effect that contextual factors, as moderating variables, have on emerging network structures. This question is explored in an academic setting, where scientific collaboration networks of Dutch Computer Science researchers are investigated (Chapter 2). Context was identified through the identification of a number of policies that attempted to steer publication within nine Dutch Computer Science departments. Scientific collaboration structures were measured through publication data, inferring collaboration through co-authorship. The success of the realized publications of the individual researchers was measured through raw delayed citation scores. I addressed this question in separate empirical studies presented in three chapters.

Three key findings emerged from this research:

(1) The policies of a professional tenure system and a publication target list within a department enhance the success of cohesive scientific collaboration networks of researchers (Chapter 3).

(2) The existence of a tenure system influences the generative roots that explain the emergence of these scientific collaboration network structures of a department (Chapter 4).

(3) Tendencies to collaborate can be explained by a common university affiliation (Chapter 5).

In Chapter 3 I investigated the effect of two departmental policies on the success of the individual network structures of Dutch Computer Science researchers. Scholars have long debated the benefits of constraint in networks the extent to which an actor's alters are connected to each other; however, optimal network structures remain debated (Burt, 2005; Coleman, 1988; Granovetter, 1985). In an attempt to shed light on this puzzle, some studies have suggested that specific network structures provide benefits depending on the context of the networks

(Battilana and Casciaro, 2012; Burt, 2000). Formal organizations, such as universities, often impose restrictions and create opportunities through policies that seek to guide behavior and outcomes. In an attempt to elucidate the role of these policies in influencing network structures and subsequent outcomes, I considered the following: which organizational policies influence networks such that individual performance improves?

This study provides clear support for a theory that stipulates the interaction of organizational policies on the success of network structures. Findings suggest that, depending on departmental policies, constraint plays different roles in the success of individual researchers' scientific collaboration networks. The policies of a professional tenure system and a target list within a department enhance the success of cohesive ego networks scientific collaboration. These policies serve as a sort of herding mechanism, working to facilitate specific types of networks of researchers. These are distinct where success in a network with low constraint allows the researcher to bridge or manage a number of clusters of collaborators, compared to success in a network that is more cohesive (high constraint), where communication barriers are reduced. These findings contribute to the ongoing debate on constraint between the efficiency of structural holes and cohesive networks, through specifying the conditions under which specific networks are successful. This research confirms previous studies that argued that high constraint in networks is most effective for completing complex knowledge tasks.

In Chapter 4 I investigated the emergence of collaboration structures using a dynamic network model to compare the mechanisms within the departments. Current methods were outlined and I proposed an extension to the use of SIENA (Snijders et al., 2010), an actor-based simulation network model, to compare the effect of different contextual factors on network dynamics. I proposed to compare models in classifying how a set of contextual conditions as defined by the bounded network relates to roles of significant generative mechanisms. Findings showed the existence of these effects and the strength of the effects vary between the departments. Linking these distinctions back to the specific classifications of exogenous contextual factors I found patterns in the specific factors in particular departments with clear policies for attempting to steer outputs by way of a formalized tenure system and a publication list generated networks with tendencies for transitive ties as well as a negative likelihood to collaborate with researchers that occupied broker positions. Thus, a context with a policies such as professional tenure system and a publication target list display structures of a networks where researchers work in close collaboration. Thus, context explains the local social mechanisms that researchers undertake to undertake collaboration.

In the final empirical chapter, Chapter 5, I investigated the entire field of Dutch Computer Science to identify a possible effect from the institutional level in a mean-field model. Scalability issues were overcome in previous models through aggregation of individual nodes. Parameters are developed using a data-aware approach which combines empirical research from Social Science and standard inferential statistics to develop a population-specific model for exploring the dynamics of collaboration in science. Thus here the context is no longer a control or a boundary condition but an explanatory factor in explaining dynamics. Findings show that past collaborative partners of ones institution plays a key role in how future collaborations unfold. With every publication with another institution the chance of collaborating with someone from the same institution increases.

In invoking the mean field model, a number of theoretical assumptions emerge. This is the theory of the field. Field theory itself is not new; Bourdieu (1998) wrote extensively on the concept, as understood within the social sciences. Exogenous factors limits the nature of an elements/actors behavior. Thus, in specifying how and when these shifts in behavior occur, we can identify the conditions under which the state of an element changes. Inversely, through the identification of contextual factors, we can identify the set of possible behaviors. Thus, elements that potentially shift have particular attributes that make them susceptible to the field effect. Thus further application of the mean field model needs to explore the implications for network theory in identifying the exogenous context factor as having a critical interplay as an explanatory variable in explaining dynamics. This research is a first step in that call.

## Implications

The context under which researchers develop and maintain their scientific collaboration networks influence their success, the behaviors employed to realize these networks and the tendency to collaborate. Within knowledge-intensive activities such as scientific collaboration, policies that attempt to avoid uncertainty alter the way in which researchers realize publications via potential collaborators. When a department has an evaluation list and a tenure system, it provides certainty about evaluation criteria. I suggest that this leads to distinct networking strategies where those in departments with specifications focus on their local position and invest in cohesive teams; whereas those without a list or tenure track system, performance criteria are uncertain and researchers must position themselves more globally. Strategic position in one context requires a different network than in the other. Researchers are facilitated and/or restrained by these specific policies about output and thus strategize accordingly. A mix of policies yields, overall, lower performance than when policies are explicit, which perhaps identifies a type of uncertainty within the context and leads to more experimentation with potential co-authors. This evidence has implications not only for how we understand dynamics but also for identifying the conditions under which individuals can employ their networks in a specific manner towards a number of possible outcomes.

These findings also contribute to our knowledge about the specific effects of different contextual factors on network structures. They provide suggestive evidence about how departments and/or universities steer success not necessarily through physical or monetary resources but through policies that promote specific behaviors. Thus it is not necessarily that top ranked universities or department attract better researchers (Allison and Long, 1990), but rather that in order to remain within the department, given the stipulations, researchers employ specific behaviors to achieve collaboration; those that do not behave this way are effectively removed or leave. Resources of top-ranked academic units serve as a selection system for cultivating a specific end. If this serves as the case, then when we consider a more global view on collaboration, we should observe that a common affiliation plays a key role in how collaboration unfolds. Given that context can be seen as a dependency for social mechanisms, a number of implications can be garnered for current theories on network dynamics. Overall it brings to light the necessary further theoretical exploration of exogenous contextual factors in constraining and/or facilitating the emergence of other roots and ultimately structures.

Further, as proved in this research, the use of the mixed-method model was an appropriate, necessary and valid choice to explore various possible effects of context. Such a toolbox allowed me to provide a broad set of evidence at which to understand the potential effects of context. This study also contributes to growing knowledge on the development of mixed methods for studies investigating both context as an antecedent as a determinant to the emergence and success of social network structures in studying large social networks. Thus, it suggests that effects need to be investigated at multiple levels  the ego, the bounded, relatively small, organizational context most often explored in social science studies of networks and a more global field approach. This allows the exploration of the complete reach of effects in explaining network antecedents.

## Policy Implications

Thus, organizational context is an important enabling or constraining condition in both the emergence of specific structures and the success of different network structures. The results from the current study suggest that knowledge intensive organizations (e.g., universities) have the means via policies to affect the success of collaboration networks among scientists. The findings that suggest that departments with both a tenure system and a publication list enhance the success of cohesive scientific collaboration networks. These policies outline a set of specifications that guide the behavior that researchers evoke to achieve publication through collaboration. This leads to distinct networking strategies where researchers in departments with specifications can focus on their local positions (i.e., within the department) and invest in cohesive teams, whereas those without a list or tenure track system, or with uncertain criteria, position themselves more globally.

These results are particularly important for individual scientists and universities. First, individual scientists can try to build their collaboration networks in accordance with the departmental policies. A first step in doing this is being aware of those you collaborate with, and also how you collaborate with others to realize publications. If you realize your publications with a core team, work to facilitate their communication and be sure to insure injections of new knowledge to maximize knowledge exchange and success through publication. If you function as a broker between different groups of coauthors, that do not collaborate with each other through publication, be aware of how you share knowledge with the different groups, and reflect on your position given different constraints of projects.

Second, universities can build on the insights of this research by streamlining their policies for collaboration and evaluation. For example, if a department has a publication list and a tenure system, it might pay off to encourage ties between collaborators by building opportunities for these collaborators to meet more often. Conversely, if a department has neither publication list nor tenure system, it might be best for performance of individual researchers to promote and encourage contacts with new collaborators, through e.g. facilitating conference visits. Concluding, both universities and researcher will profit from more attention to the contingencies of the social networks that they are embedded.

The success of cohesive networks in departments with publication steering policies are vulnerable to the increasing scarcity and uncertainty of tenured research positions. This is a threat to the overall quality of both a discipline and science system. The effect of these policies come into question when the incentives of promotion from successful publications become limited or even nearly impossible. Policymakers should adhere to the importance of balancing relatively cheap, junior research positions through funding mechanisms, departmental reorganizations and the like where tenure becomes only achievable for the very few; with funding for longer term research projects, and permanent chairs to sustain a healthy system.

In addition, in regards to Dutch Computer Science as a field. Despite that conference proceedings comprise the largest share of publications and are also among the highest cited publications within computer science, they still remain often absent from commonly used indexes and formal reviews and evaluations by faculties and universities. This distinct communication practice is often misinterpreted as a way to publish low quality or findings of less theoretical or practical value; when in reality these proceedings are significant products that serve both as markers of the patterns of collaboration that researchers undertake to produce knowledge, as well as the attributed quality of knowledge. Evaluators should take note of the impact of evaluated knowledge outputs of multiple forms and the networks that researchers employ in developing scientific knowledge when considering candidates for positions, funding and reviewing quality.

## Conclusion

The context under which researchers develop and maintain their scientific collaboration networks influence their success, the behaviors employed to realize these networks and the tendency to collaborate. This evidence has implications not only for how we understand dynamics but also for identifying the conditions under which individuals can employ their networks in a specific manner towards a number of possible outcomes. This sheds new light on previous network research, in particular the large body of research that has solely selected networks as a boundary condition for investigating networking mechanisms. This is just the first evidence to suggest context as a determinant and as described in the discussion deserves further theoretical exploration for confirming these effects. Although, if it holds true it puts to question the validity of a number of findings. Particularly those of studies where mechanisms attributed singly to individuals are said to be the driving force we must consider the conditions under which these local mechanisms arose. Thus suggesting different *modi operandi* for explaining the emergence and success of network structures.

This research has made a contribution to theoretical, methodological and practical knowledge for understanding the role of context in network dynamics. This research is among the first to provide empirical evidence within the framework of structuration to consider how exogenous contextual factors act as determinants. The exploration of a mean field model, as well as the implementation of traditional social network analysis techniques, proved that context has an effect on networking behavior. Thus, in understanding social network dynamics, context truly matters.

## Nederlandse Samenvatting

Sociale netwerken zijn van belang. Kenmerken van sociale netwerken hebben gevolgen op uitkomsten (Burt, 2005; Coleman, 1988; Fleming et al., 2007). Minder begrepen zijn de precieze mechanismen die leiden tot netwerkstructuren. Om de dynamiek van sociale netwerken te begrijpen, zijn de mechanismen waardoor netwerken ontstaan onderzocht. Zo wordt meer inzicht verkregen in hoe dit proces werkt.

In dit proefschrift is een aantal generatieve mechanismen gedentificeerd als antecedenten die leiden tot netwerkstructuren. Deze generatieve mechanismen kennen hun oorsprong in drie roots:

(1) alleen-netwerk factoren;

(2) individuele factoren; en

(3) contextuele factoren.

Deze roots beschrijven de oorsprong van deze netwerkmechanismen. De rol die deze generatieve roots spelen bij de ontwikkeling van netwerkstructuren blijven onderwerp van discussie, theoriën suggereren verschillende *modi operandi* (Ahuja et al., 2012; Liben-Nowell and Kleinberg, 2007; Shumate and Contractor, 2013; Whitbred et al., 2011).

Ik stel dat de context over het hoofd wordt gezien, ondanks het toenemende bewijs dat het succes van verschillende netwerkstructuren afhankelijk is van verschillende organisatorische processen (Battilana and Casciaro, 2012; Burt, 2000). Door de invloed die contextuele factoren hebben op de uitwerking van netwerkstructuren te verklaren, vergroot ik met dit onderzoek het inzicht in de modi operandi van netwerken. Om de rol van context op de dynamiek sociale netwerk te verkennen, heb ik de netwerkstructuren vanuit drie invalshoeken bestudeerd:

(1) het succes van individuele netwerkstructuren (hoofdstuk 3);

(2) de patronen van generatieve mechanismen in netwerken (hoofdstuk 4); en

(3) de tendensen van interacties (hoofdstuk 5).

Door middel van literatuuronderzoek laat ik zien dat er geen bestaand kader is dat het samenspel van contextuele factoren en individuele factoren van netwerkstructuren verklaart. Ik heb daarom een het theoretisch kader van structuratietheorie toegepast om de huidige kennis over de rol die generatieve roots spelen in de opkomst van netwerken te conceptualiseren. Ik stel voor om de structuratietheorie te gebruiken om de rol van de context op het netwerkstructuren af te bakenen en ik schets een theoretisch kader gebaseerd op het concept van *duality*.

*Duality* is de wisselwerking tussen rules en resources die uitleggen hoe structuren in dienst van individuen gebruikt worden om een actie te realiseren (Giddens, 1984). Ik heb rules als contextuele factoren beschouwd die gedentificeerd worden door middel van beleid als samenspel van resources - endogene netwerk factoren en individuele kenmerken gemeten vanaf eerdere netwerkposities en eigenschappen van de individuele onderzoeker - om deze verschillen in netwerkstructuren te beschrijven.

Bij het nastreven van deze verkenning naar de exogene invloed op de dynamiek waren verschillende methodologische uitbreidingen nodig. Ik heb een mixed-methods netwerkbenadering gebruikt om het effect te evalueren dat contextuele factoren hebben op de opkomende netwerkstructuren. Ik heb deze vraag verkend in een academische setting, waarin de wetenschappelijke samenwerkingsnetwerken van Nederlandse Informaticaonderzoekers zijn onderzocht (hoofdstuk 2). Door de identificatie van een aantal beleidsmaatregelen die probeerden om publicaties te sturen binnen negen Nederlandse Informatica afdelingen, kon context worden gedentificeerd. Wetenschappelijke samenwerkingsstructuren zijn gemeten door middel van gegevens over publicaties, samenwerking is door middel van co-auteurschap afgeleid. Het succes van de gerealiseerde publicaties van de individuele onderzoekers werd gemeten door middel van *raw-delayed* citatiescores. Ik heb deze vraag in afzonderlijke empirische studies gepresenteerd in drie hoofdstukken. Drie belangrijkste bevindingen die uit dit onderzoek zijn voortgekomen:

(1) Het beleid voor een professionele tenure-track systeem en een publicatie targetlijst binnen een afdeling vergroten het succes van *cohesive* wetenschappelijke samenwerkingsnetwerken van onderzoekers (hoofdstuk 3).

(2) Het bestaan van een tenure-track systeem beïnvloedt de generatieve roots die de opkomst van deze wetenschappelijke samenwerkingsnetwerkstructuren van een afdeling verklaren(hoofdstuk 4).

(3) De geneigdheid om samen te werken kan worden verklaard door een gemeenschappelijke band met een universiteit (hoofdstuk 5).

In hoofdstuk 3 heb ik het effect van twee vormen van beleid van departementen onderzocht op het succes van de individuele netwerkstructuren van Nederlandse Informaticaonderzoekers. Wetenschappers hebben lang gediscussieerd over de voordelen van constraint in netwerken– de mate waarin alters van een actor met elkaar zijn verbonden; optimale netwerkstructuren blijven echter een onderwerp van debat. In een poging duidelijkheid te scheppen in deze puzzel, hebben enkele studies gesuggereerd dat specifieke netwerkstructuren voordelen bieden, afhankelijk van de context van de netwerken. Formele organisaties, zoals universiteiten, leggen vaak beperkingen op en creren kansen door middel van beleid dat er op is gericht om gedrag en resultaten te sturen. In een poging om de rol van dit beleid op het beïnvloeden van netwerkstructuren en de daaropvolgende gevolgen te verklaren, heb ik de volgende vraag gesteld: welk organisatorisch beleid beïnvloeden netwerken zodaning dat de individuele prestaties verbeteren?

Dit onderzoek biedt duidelijke steun voor een theorie die de interactie van organisatorisch beleid op het succes van netwerkstructuren bepaalt. Bevindingen suggereren dat, afhankelijk van het departementaal beleid, *constraint* uiteenlopende rollen speelt voor het succes van de wetenschappelijke samenwerkingsnetwerken van individuele onderzoekers. Beleid voor een professionele tenure-track systeem en een publicatie targetlijst binnen een afdeling vergroot het succes van wetenschappelijke samenwerking binnen *cohesive* ego netwerken. Dit beleid dient als een mechanisme dat de samenwerking van specifieke vormen van netwerken van onderzoekers vergemakkelijkt. Deze zijn duidelijk te onderscheiden daar waar succes in een netwerk met *low constraint* de onderzoeker in staat stelt cluster van medewerkers te managen of te verbinden, in vergelijking tot het succes van een netwerk dat meer samenhang vertoont (*high constraint*), waar de communicatiebarrières worden verminderd. Deze bevindingen dragen bij aan het lopende debat over de *constraint* tussen de efficiëntie van *structural holes* en *cohesive* netwerken (Burt, 2005; Coleman, 1988; Granovetter, 1985), door vermelding van de voorwaarden waaronder specifieke netwerken succesvol zijn. Dit onderzoek bevestigt de uitkomsten van eerdere studies die aangetoond hebben dat *high constraint* in netwerken het meest effectiëf zijn voor het voltooien van complexe kennistaken.

In hoofdstuk 4 onderzocht ik de opkomst van samenwerkingsstructuren met behulp van een dynamisch netwerkmodel om de mechanismen binnen de afdelingen te vergelijken. Ik schetste de huidige methodes en stel een uitbreiding voor van het gebruik van SIENA (Snijders et al., 2010), een *actor based simulation* netwerk model, om het effect van verschillende factoren op contextuele netwerkdynamiek te vergelijken. Ik stel voor om modellen te vergelijken door middel van een set van geclassificeerde contextuele condities die gedefinieerd worden door de *bounded* netwerken in relatie tot de rollen van significante generatieve mechanismen. Bevindingen suggereren dat de aanwezigheid van deze effecten en de sterkte van de effecten verschillen tussen afdelingen. Door deze verschillen terug te koppelen naar de specifieke classificaties van exogene contextuele factoren, vinden we patronen in deze specifieke factoren. Deze wijzen erop dat afdelingen die een duidelijk beleid hebben om de output te sturen, door middel van een geformaliseerd tenure-track systeem en een publicatie targetlijst, netwerken genereren met tendensen voor *transitive ties*. Daarbij bestaat hier een negatieve kans van samenwerking met onderzoekers die *broker positions* bezetten. Derhalve kan geconcludeerd worden dat een context met een beleid voor professionele tenure-track systeem en een publicatie targetlijst een structuur biedt van een netwerk waar onderzoekers nauw samen werken. Context verklaart zo de lokale sociale mechanismen die onderzoekers ondernemen om samenwerking mogelijk te maken.

In het laatste empirische hoofdstuk, hoofdstuk 5, is het gehele vakgebied van de Nederlandse Informatica onderzocht om een mogelijk effect te identificeren ophet institutioneel niveau door middel van het *mean field model*. Wij zijn in staat om *scalability issues* te overkomen die zich voordoen in voorgaande modellen door middel van de samenvoeging van afzonderlijke nodes. Met behulp van een *data-aware* aanpak die empirisch onderzoek van sociale wetenschappen en standaard inferentile statistiek combineert zijn parameters ontwikkeld om een populatie-specifiek model te ontwikkelen om de dynamiek van samenwerking in de wetenschap te verkennen. Context is zo niet langer een controle of een randvoorwaarde, maar een verklarende factor in de netwerk dynamiek. Bevindingen tonen aan dat samenwerkingspartners uit het verleden van iemands instelling een belangrijke rol speelt in hoe de toekomstige samenwerkingen zich ontwikkeld. Bij elke publicatie met een andere instelling wordt de kans op samenwerking met iemand van diezelfde instelling groter.

In het gebruik van het *mean field model* zijn een aantal theoretische aannames ontstaan. Dit is de theorie van het veld. Veldtheorie zelf is niet nieuw; Bourdieu (1998) schreef uitgebreid over het concept, zoals begrepen binnen de sociale wetenschappen. Exogene factoren beperken de aard van het gedrag van een element/actor. Door het specificeren van hoe en wanneer deze verschuivingen in gedrag optreden, kunnen we dus aangeven wat de voorwaarden zijn waaronder de toestand van een element verandert. Omgekeerd kunnen we de set van mogelijke gedragingen bloot leggen door de identificatie van contextuele factoren. Elementen die mogelijk andere gedrag gaan vertonen, hebben kenmerken die hen gevoelig maken voor het field effect. Verdere toepassingen van het *mean field model* moeten onderzoeken wat de implicaties zijn voor netwerktheorie in haar mogelijkheid om de exogene contextuele factor te identificeren, met een kritische interactie als verklarende variabele in de te verklaren dynamiek. Dit onderzoek is een eerste stap op deze weg.

## Implicaties

De context waarin onderzoekers hun wetenschappelijke samenwerkingsnetwerken ontwikkelen en in stand houden, beïnvloedt hun succes en hun acties om deze netwerken tot stand te brengen en hun neiging om samen te werken. Binnen kennisintensieve activiteiten, zoals wetenschappelijke samenwerking, verandert beleid, dat als doel heeft onzekerheid weg te nemen, de manier waarop onderzoekers publicaties tot stand worden gebracht via mogelijke samenwerking. Als een afdeling een publicatie targetlijst en een tenure-track systeem heeft, biedt dat zekerheid en duidelijkheid over de manier waarop de onderzoeker beoordeeld wordt.

Ik stel dat dit leidt tot verschillende netwerkstrategieën. Waar diegenen in afdelingen met specificaties zich richten op hun lokale positie en investeren in *cohesive* teams, zullen de onderzoekers zonder een tenure-track systeem of publicatie targetlijst en met onduidelijke prestatiecriteria zich meer globaal moeten positioneren. Onderzoekers worden gefaciliteerd en/of beperkt door dit specifieke beleid over output en kiezen dienovereenkomstig hun strategie. Een mix van beleidsdoelen leidt in het algemeen tot lagere prestaties dan wanneer beleid expliciet is; dit genereert onzekerheid binnen de context en leidt tot meer experimenten met potentile co-auteurs. Dit bewijs heeft niet alleen gevolgen voor de manier waarop we de dynamiek begrijpen, maar ook voor het vaststellen van de voorwaarden waaronder individuen hun netwerken kunnen toepassen op een specifieke manier, om te komen tot een aantal mogelijke uitkomsten.

Deze bevindingen dragen ook bij aan onze kennis over de specifieke effecten van verschillende contextuele factoren op netwerkstructuren. Ze bieden suggestief bewijs over hoe afdelingen en/of universiteiten succes sturen, niet per se door middel van fysieke of monetaire middelen, maar door middel van beleid dat specifiek gedrag bevordert. Het is dus niet per se zo dat de top gerangschikte universiteiten of afdeling betere onderzoekers aantrekken (Allison en Long, 2012), maar eerder dat om binnen de afdeling te blijven, gezien de omstandigheden, onderzoekers specifiek gedrag vertonen om samenwerking te bereiken; degenen die niet zulke acties ondernemen worden ontslagen of vertrekken. Middelen van top gerangschikte academische eenheden dienen als een selectiemechanisme voor het cultiveren van een specifiek doel. Als dit het geval is, als we een meer globale visie hebben op samenwerking, moeten we constateren dat een common affiliatie een belangrijke rol speelt in de manier waarop de samenwerking zich ontvouwt. Gezien het feit dat context kan worden gezien als een afhankelijkheid van sociale mechanismen, kunnen een aantal implicaties worden vergaard voor de huidige theorien over netwerkdynamiek. Over het algemeen brengt dit onderzoek de nodige verdere theoretische verkenning van exogene contextuele factoren aan het licht en hun rol in de beperking danwel in de opkomst van andere roots en uiteindelijk structuren.

Dit onderzoek toont verder aan dat het gebruik van een mixed-methode model een passende, noodzakelijke en geldige keuze is om de verschillende mogelijke effecten van context te verkennen. Een dergelijke toolbox stelde mij in staat om een brede set van bewijsmateriaal te verzamelen, waarmee ik de mogelijke effecten van de context kon begrijpen. Deze studie draagt eveneens bij aan de groeiende kennis over de ontwikkeling van mixed methods voor studies naar zowel context als antecedent als een bepalende factorvoor de opkomst en het succes van sociale netwerkstructuren en het bestuderen van grote sociale netwerken. Het onderzoek suggereert dat de effecten moeten worden onderzocht op verschillende niveaus: het ego, de bounded, relatief kleine, organisatorische context die het meest wordt onderzocht in sociaal-wetenschappelijk onderzoek van netwerken, en een meer globale veld aanpak. Hierdoor het volledige bereik van de effecten worden verkend bij het verklaren van het netwerk antecedenten.

## Beleidsimplicaties

Organisatiecontext speelt een belangrijke rol bij het mogelijk maken of bemoeilijken van de totstandkoming van specifieke structuren en het succes van verschillende netwerkstructuren. De resultaten van het huidige onderzoek duiden erop dat kennisintensieve organisaties (bijvoorbeeld universiteiten) over de middelen beschikken om via het beleid het succes van samenwerkingsverbanden tussen wetenschappers te beïnvloeden. De bevindingen tonen verder dat afdelingen met zowel een tenure-track systeem als een publicatie targetlijst het succes van wetenschappelijke samenwerkingsverbanden vergroten. Dit beleid schetst een aantal leidraden dat het gedrag van de onderzoeker stuurt om publicaties tot stand te brengen door middel van samenwerking. Dit leidt tot verschillende netwerkstrategieën waar onderzoekers binnen de afdelingen met specificaties zich kunnen richten op hun lokale posities (dat wil zeggen, binnen de afdeling) en investeren in hechte teams, terwijl degenen zonder een lijst of tenure-track systeem, of met onduidelijke criteria, zich meer globaal positioneren.

Deze resultaten zijn in het bijzonder belangrijk voor individuele wetenschappers en universiteiten. Ten eerste kunnen individuele wetenschappers proberen om hun samenwerkingsnetwerken op te bouwen in overeenstemming met het afdelingsbeleid. Een eerste stap om dit te doen is je bewust te zijn van de mensen met wie je samenwerkt, en ook hoe je met anderen samenwerkt om publicaties te realiseren. Als je publicaties realiseert met een kernteam, is het van belang om hun communicatie te vergemakkelijken en zorgt voor een instroom van nieuwe infromate om de kennisuitwisseling te vergroten en zodoende het succes door publicaties te realiseren. Als je functioneert als een makelaar tussen verschillende groepen coauteurs, die niet met elkaar samenwerken door middel van publicaties, moet je er bewust van zijn hoe je de kennis met de verschillende groepen deelt, en nadenken over je positie gezien de verschillende beperkingen van de projecten.

Ten tweede kunnen universiteiten bouwen op de inzichten van dit onderzoek door het stroomlijnen van hun beleid voor samenwerking en evaluatie. Als een afdeling bijvoorbeeld een publicatie targetlijst en een tenure-track systeem heeft, kan het lonend zijn om de banden tussen de medewerkers te stimuleren door de kansen te vergroten dat deze medewerkers elkaar vaker ontmoeten. Omgekeerd, als een afdeling noch een publicatie targetlijst noch een tenure-track systeem heeft, is het misschien het beste om prestaties van individuele onderzoekers te bevorderen en contacten met nieuwe medewerkers aan te moedigen, bijvoorbeeld door middel van het faciliteren van conferentiebezoeken. Kortom, beide universiteiten en onderzoeker zullen profiteren van meer aandacht voor de contingencies van de sociale netwerken waarin ze zijn ingebed.

Het succes van *cohesive* netwerken in afdelingen met een publicatiebeleid wordt bedreigd door de afnemende beschikbaarheid van tenure-track posities, zogenaamde vaste aanstellingsonderzoeksposities. Dit is een bedreiging voor zowel de algehele kwaliteit van de discipline en het wetenschapsysteem. Het effect van dit beleid wordt in twijfel getrokken, zonder dat daar een positieve prikkel tegen over staat voor het behalen van succesvolle publicaties. Beleidsmakers moeten zich bewust zijn van het belang van een goede balans tussen relatief goedkope junior onderzoeksposities door middel van financieringsmechanismen,departementale reorganisaties en dergelijke, waar een tenure-track positie alleen haalbaar is voor een hele kleine groep onderzoekers met lange termijn financiering en permanente leerstoelen om een gezond systeem in stand te houden.

Bovendien, met betrekking tot Nederlandse Informatica als een veld. Ondanks dat de de publicaties voor het grootste deel bestaat uit conferentie proceedings en teven behoren tot de hoogst geciteerde publicaties binnen de informatica, zijn ze nog steeds geen onderdeel van de veelgebruikte indexen en formele beoordelingen en evaluaties door faculteiten en universiteiten. Deze manier van publiceren (conference proceedings) wordt in de praktijk vaak gezien als een manier om lage kwaliteitsonderzoek, of bevindingen van minder theoretische of praktische waarde, te publiceren. In werkelijkheid zijn deze proceedings belangrijke publicaties die zowel de patronen van samenwerking die wetenschappers ondernemen weergeven om kennis te produceren alsmede de toegekende waarde door citatiescores. Evaluatoren moeten zich bewust zijn van de impact van de kennis output van meerde types van publicaties bij het overwegen van kandidaten voor functies, financieringen en beoordelen van de kwaliteit.

## Conclusie

De context waarin onderzoekers zich ontwikkelen en de mate waarin zij hun wetenschappelijke samenwerkingsnetwerken onderhouden beïnvloedt hun succes, het toegepaste gedrag om deze netwerken te realiseren en de neiging om samen te werken. Dit bewijs heeft niet alleen gevolgen voor de manier waarop we de dynamiek begrijpen, maar ook voor het vaststellen van de voorwaarden waaronder individuen hun netwerken op een specifieke manier kunnen inzetten om *outcomes*te genereren. Dit werpt nieuw licht op voorgaand netwerkonderzoek, voornamelijk de grote hoeveelheid onderzoek dat netwerkmechanismen uitsluitend selecteert als een randvoorwaarde. Dit onderzoek is slechts het eerste onderzoek dat suggereert dat context een bepalende factor is en zoals beschreven in de discussie verdient dit nadere theoretische exploratie om deze effecten te bevestigen.

Als dit wordt bevestigd, zal het de validiteit van meerdere onderzoeksuitkomsten in twijfel trekken. In het bijzonder bij die onderzoeken waar mechanismen uitgelegd worden aan de hand van individueel gedrag, zal gekeken moeten worden naar de context waarin deze mechanismen optraden. Dit stelt ons in staat om verschillende *modi operandi* te gebruiken om de opkomst en succes van netwerkstructuren te verklaren.

Dit onderzoek heeft een bijdrage geleverd aan theoretische, methodologische en praktische kennis voor het begrijpen van de rol die context speelt in netwerkdynamiek. Dit onderzoek is een van de eerste onderzoeken dat binnen de structuratietheore empirisch bewijs levert over hoe exogene contextuele factoren een rol als determinant spelen. De verkenning van een *mean field* model, als ook de toepassing van traditionele sociale netwerkanalyse technieken hebben aangetoond dat context een effect heeft op netwerkgedrag. Om sociale netwerk dynamiek te begrijpen, is context dus uitermate belangrijk.

# Bibliography


3TU (2014). http://www.3tu.nl.

Adamic, L. A. and Huberman, B. A. (2000). Power-law distribution of the world wide web. *Science*, 287(5461):2115–2115.

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, 45(3):425–455.

Ahuja, G., Soda, G., and Zaheer, A. (2012). The genesis and dynamics of organizational networks. *Organization Science*, 23:434–448.

Akkermans, H. (2012). Web dynamics as a random walk: How and why power laws occur. In *Proceedings of the 3rd Annual ACM Web Science Conference*, 1–10, New York, NY. ACM.

Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97.

Allison, P. D. and Long, J. S. (1990). Departmental Effects on Scientific Productivity. *American Sociological Review*, 55(4):469–478.

Anderson, C. J., Wasserman, S., and Crouch, B. (1999). A P* Primer: Logit Models for Social Networks. *Social Networks*, 21(1):37–66.

ArnetMiner (2012). http://arnetminer.org/.

ASCI (2014). http://www.asci.tudelft.nl/.

Bakhshi, R., Endrullis, J., Endrullis, S., Fokkink, W., and Haverkort, B. (2010). Automating the mean-field method for large dynamic gossip networks. In *Proceedings of Quantitative Evaluation of Systems*, 241–250. New York, NY: IEEE Computer Society.

Balkundi, P., Kilduff, M., Barsness, Z. I., and Michael, J. H. (2007). Demographic antecedents and performance consequences of structural holes in work teams. *Journal of Organizational Behavior*, 28(2):241–260.

Bar-Ilan, J. (2010a). Citations to the "introduction to informetrics" indexed by WOS, scopus and google scholar. *Scientometrics*, 82(3):495–506.

Bar-Ilan, J. (2010b). Web of science with the conference proceedings citation indexes: the case of computer science. *Scientometrics*, 83(3):809–824.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Barabási, A. L., Jeong, H., Nda, Z., Ravasz, E., Schubert, A., and Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3):590–614.

Barley, S. R. and Tolbert, P. S. (1997). Institutionalization and structuration: Studying the links between action and institution. *Organization Studies*, 18(1):93–117.

Battilana, J. and Casciaro, T. (2012). Change agents, networks, and institutions: A contingency theory of organizational change. *Academy of Management Journal*, 55(2):381–398.

Birkholz, J. M., Bakhshi, R., Harige, R., van Steen, M., and Groenewegen, P. (2012). Scalable Analysis for Large Social Networks: The Data-Aware Mean-Field Approach. In *Social Informatics*, volume 7710 of *LNCS*, 406–419. Berlin Heidelberg, Germany: Springer.

Birkholz, J. M., Wang, S., Groth, P., and Magliacane, S. (2013). Who Are We Talking About? Identifying Scientific Populations Online. In *Semantic Web and Web Science*, 237–250. Berlin Heidelberg, Germany: Springer.

Bishop, C. M. (1995). *Neural Networks for Pattern Recognition*. Oxford, England: Oxford University Press.

Blau, P. (1977). *Inequality and Heterogeneity*. New York, NY: The Free Press.

Blau, P. M. (1964). *Exchange and power in social life*. Piscataway, NJ: Transaction Publishers.

Blau, P. M. (1974). *On the nature of organizations*. New York City, New York: Wiley New York.

Boardman, P. and Corley, E. (2008). University research centers and the composition of research collaborations? *Research Policy*, 37(5):900–913.

Boardman, P. C. and Ponomariov, B. L. (2007). Reward Systems and NSF University Research Centers: The Impact of Tenure on University Scientists' Valuation of Applied and Commercially Relevant Research. *The Journal of Higher Education*, 78(1):51–70.

Borgatti, S., Mehra, A., Brass, D., and Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323(5916):892–895.

Borgatti, S. P. and Cross, R. (2003). A relational view of information seeking and learning in social networks. *Management Science*, 49(4):432–445.

Borgatti, S. P., Everett, M. G., and Freeman, L. C. (2002). *Ucinet for Windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.

Borgatti, S. P. and Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. *Journal of Management*, 29(6):991–1013.

Bourdieu, P. (1998). *Practical reason: On the theory of action*. Palo Alto, CA: Stanford University Press.

Bozeman, B. and Corley, E. (2004). Scientists' collaboration strategies: Implications for scientific and technical human capital. *Research Policy*, 33(4):599–616.

Bozeman, B. and Gaughan, M. (2007). Impacts of grants and contracts on academic researchers interactions with industry. *Research Policy*, 36(5):694–707.

Bozeman, B. and Gaughan, M. (2011). How do men and women differ in research collaborations? An analysis of the collaborative motives and strategies of academic researchers. *Research Policy*, 40(10):1393–1402.

Brass, D. J. (1984). Being in the right place: A structural analysis of individual influence in an organization. *Administrative Science Quarterly*, 29:518–539.

Burkhardt, M. E. and Brass, D. J. (1990). Changing patterns or patterns of change: The effects of a change in technology on social network structure and power. *Administrative Science Quarterly*, 35(1):104–127.

Burns, T. E. and Stalker, G. M. (1961). *The management of innovation*. University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship, Urbana-Champaign, IL.

Burt, R. S. (1980). Models of network structure. *Annual Review of Sociology*, 79–141.

Burt, R. S. (1992). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.

Burt, R. S. (2000). The network structure of social capital. *Research in Organizational Behavior*, 22:345–423.

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, 110(2):349–399.

Burt, R. S. (2005). *Brokerage and closure: An introduction to social capital*. Cambrige, MA: Oxford University Press.

Burt, R. S., Hogarth, R. M., and Michaud, C. (2000). The social capital of french and american managers. *Organization Science*, 11(2):123–147.

Burt, R. S., Jannotta, J. E., and Mahoney, J. T. (1998). Personality correlates of structural holes. *Social Networks*, 20(1):63–87.

Carley, K. M. (1997). Extracting team mental models through textual analysis: Summary. *Journal of Organizational Behavior*, 18:533–558.

Carley, K. M. and Palmquist, M. (1992). Extracting, Representing, and Analyzing Mental Models. *Social Forces*, 70(3):601–636.

Cartwright, D. and Harary, F. (1956). Structural balance: a generalization of Heider's theory. *Psychological Review*, 63(5):277–293.

Cole, J. R. and Zuckerman, H. (1984). The productivity puzzle: Persistence and change in patterns of publication of men and women scientists. *Advances in Motivation and Achievement*, 2:217–258.

Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94:S95–S120.

Corley, E. A., Boardman, P. C., and Bozeman, B. (2006). Design and the management of multi-institutional research collaborations: Theoretical implications from two case studies. *Research Policy*, 35(7):975–993.

Covaleski, M. A. and Dirsmith, M. W. (1988). An institutional perspective on the rise, social transformation, and fall of a university budget category. *Administrative Science Quarterly*, 562–587.

Crane, D. (1969). Social structure in a group of scientists: A test of the" invisible college" hypothesis. *American Sociological Review*, 34(3):335–352.

CTIT (2014). `http://www.utwente.nl/ctit/`.

Cummings, J. N. and Cross, R. (2003). Structural properties of work groups and their consequences for performance. *Social Networks*, 25(3):197–210.

Daft, R. L. (1982). Bureaucratic versus nonbureaucratic structure and the process of innovation and change. *Research in the Sociology of Organizations*, 1:129–166.

Danowski, J. (1993). Network analysis of message content. In Richards, W. and Barnett, G., editors, *Progress in Communication Science*, chapter XII, 197–222. New York, NY: Ablex.

Danowski, J. A. and Edison-Swift, P. (1985). Crisis effects on intraorganizational computer-based communication. *Communication Research*, 12(2):251–270.

Database Systems and Logic Programming (2013). http://dblp.uni-trier.de/.

de Solla Price, D. (1965). Introduction to the special issue on network dynamics. *Science*, 149(3683):510–515.

deB Beaver, D. and Rosen, R. (1979). Studies in scientific collaboration part III. professionalization and the natural history of modern scientific co-authorship. *Scientometrics*, 1(3):231–245.

Defazio, D., Lockett, A., and Wright, M. (2009). Funding incentives, collaborative dynamics and scientific productivity: Evidence from the EU framework program. *Research Policy*, 38(2):293–305.

Denning, P. (2000). Computer science: The discipline. In Ralston, A. and Hemmendinger, D., editors, *Encyclopedia of Computer Science*, 405–419. London, England: Nature Publishing Group.

Desanctis, G. and Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science*, 5(2):121–147.

Diesner, J. and Carley, K. M. (2005). Revealing social structure from texts: meta-matrix text analysis as a novel method for network text analysis. In Narayanan, V. K. and Amstrong, D. J., editors, *Causal mapping for information systems*, chapter IV, 81–108. Hershey, PA: Idea Group Publishing.

DiMaggio, P. J. (1988). Interest and agency in institutional theory. In Zucker, L. G., editor, *Institutional Patterns and Organizations: Culture and Environment*, 3–21. Cambridge, MA: Ballinger.

DiMaggio, P. J. and Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2):147–160.

Ding, Y. (2011). Scientific collaboration and endorsement: Network analysis of coauthorship and citation networks. *Journal of Informetrics*, 5(1):187–203.

Drott, M. C. (1995). Reexamining the role of conference communication papers in scholarly. *Society*, 46(4):299–305.

Durkheim, E. (1951). *Suicide: A study in sociology (JA Spaulding & G. Simpson, trans.)*. Glencoe, IL: Free Press. (Original work published 1897).

Eden, A. H. (2007). Three paradigms of computer science. *Minds and Machines*, 17(2):135–167.

Ennett, S. and Bauman, K. (1994). The contribution of influence and selection to adolescent peer group homogeneity, the case of adolescent cigarette smoking. *Journal of Personality and Social Psychology*, 67:653–663.

European Communities (2014). The EU Information Technologies Program. http://www.cordis.europa.eu/esprit/home.html.

Fairweather, J. (2002). The mythologies of faculty productivity: Implications for institutional policy and decision making. *The Journal of Higher Education*, 73(1):26–48.

Fairweather, J. S. and Beach, A. L. (2002). Variations in faculty work at research universities: implications for state and institutional policy. *The Review of Higher Education*, 26(1):97–115.

Falk-Krzesinski, J., H., Börner, K., Contractor, N., Fiore, S. M., Hall, K. L., Keyton, J., Spring, B., Stokols, D., Trochim, W., and Uzzi, B. (2010). Advancing the science of team science. *Clinical and Translational Science*, 3(5):263–266.

Faust, K. and Skvoretz, J. (2002). Comparing networks across space and time, size and species. *Sociological Methodology*, 32(1):267–299.

Fleming, L., King III, C., and Juda, A. I. (2007). Small worlds and regional innovation. *Organization Science*, 18(6):938–954.

Foucault, M. (1977). *Discipline and punish: The birth of the prison*. New York, NY: Vintage Books.

Franceschet, M. (2010). A comparison of bibliometric indicators for computer science scholars and journals on web of science and google scholar. *Science*, 83(1):243–258.

Garg, K. and Padhi, P. (2001). A study of collaboration in laser science and technology. *Scientometrics*, 51(2):415–427.

Garud, R., Hardy, C., and Maguire, S. (2007). Institutional entrepreneurship as embedded agency: An introduction to the special issue. *Organization Studies*, 28(7):957–969.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., and Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. London, England: Sage.

Giddens, A. (1984). *The constitution of society: Outline of the theory of Structuration*. Oakland, CA: University of California Press.

Goldberg, D. S. and Roth, F. P. (2003). Assessing experimentally derived interactions in a small world. *Proceedings of the National Academy of Sciences*, 100(8):4372–4376.

Goldsman, D. and Nelson, B. L. (1998). Statistical screening, selection, and multiple comparison procedures in computer simulation. In *Proceedings of the 30th conference on Winter simulation*, 159–166. New York, NY: IEEE Computer Society Press.

Goodrum, A. A., Mccain, K. W., Lawrence, S., and Giles, C. L. (2001). Scholarly publishing in the internet age: a citation analysis of computer science literature. *Information Processing & Management*, 37:661–675.

Granovetter, M. (1973). The strength of weak ties. *American Sociological Review*, 78:1360–1380.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, 91(3):481–510.

Greene, M. (2007). The demise of the lone author. *Nature*, 450(7173):1165.

Greenwood, R., Raynard, M., Kodeih, F., Micelotta, E. R., and Lounsbury, M. (2011). Institutional complexity and organizational responses. *The Academy of Management Annals*, 5(1):317–371.

Grossman, J. W. (2005). Patterns of collaboration in mathematical research. *Notices of the AMS*, 52(1):35–41.

Guimera, R., Uzzi, B., Spiro, J., and Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, 308(5722):697–702.

Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative Science Quarterly*, 44(1):82–111.

Hansen, M. T., Podolny, J. M., and Pfeffer, J. (2001). So many ties, so little time: A task contingency perspective on corporate social capital. *Research in the Sociology of Organizations*, 18:21–57.

Hicks, D. M. and Katz, J. S. (1996). Where is science going? *Science, Technology & Human Values*, 21(4):379–406.

Hudson, J. (1996). Trends in multi-authored papers in economics. *The Journal of Economic Perspectives*, 10(3):153–158.

Ibarra, H. (1992). Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Administrative Science Quarterly*, 37:422–447.

Ibarra, H. and Andrews, S. B. (1993). Power, social influence, and sense making: Effects of network centrality and proximity on employee perceptions. *Administrative Science Quarterly*, 277–303.

Informatics Europe (2014). http://www.informatics-europe.org/.

IPA (2014). http://www2.win.tue.nl/ipa/.

Ivankova, N. V., Creswell, J. W., and Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, 18(1):3–20.

Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43.

Khanafiah, D. and Situngkir, H. (2004). Social balance theory: revisiting heider's balance theory for many agents. Technical Report WPN2004BF1, Java, Indonesia: Bandung Fe Institute.

Kilduff, M. and Krackhardt, D. (2008). *Interpersonal networks in organizations*. Cambridge, England: Cambridge University Press.

Klein, K. J., Lim, B.-C., Saltz, J. L., and Mayer, D. M. (2004). How do they get there? An examination of the antecedents of centrality in team networks. *Academy of Management Journal*, 47(6):952–963.

Knecht, A., Snijders, T. A., Baerveldt, C., Steglich, C. E., and Raub, W. (2010). Friendship and delinquency: Selection and influence processes in early adolescence. *Social Development*, 19(3):494–514.

Knoke, D. (1990). Networks of political action: Toward theory construction. *Social forces*, 68(4):1041–1063.

Knorr-Cetina, K. (1981). *The manufacture of knowledge*. Oxford, England: Pergamon Press.

Koffka, K. (1935). *Principles of Gestalt psychology*. San Diego, CA: Harcourt Brace.

Krackhardt, D. (1987). Cognitive social structures. *Social Networks*, 9(2):109–134.

Krackhardt, D. (1992). The strength of strong ties: the importance of philos in organizations. *Networks and organizations: Structure, Form, and Action*, 216:216–239.

Krackhardt, D. (1999). The ties that torture: Simmelian tie analysis in organizations. *Research in the Sociology of Organizations*, 16(1):183–210.

Krebs, V. (2002). Mapping networks of terrorist cells. *Connections*, 24(2):43–52.

Kuhn, T. S. (1996). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Kuiper, R. (2014). Zorg in de universitaire wereld: de wetenschapper als flexwerker. *Volkskrant*.

Lakatos, I. and Musgrave, A. (1970). Criticism and the growth of knowledge. *Proceedings of the International Colloquium in the Philosophy of Science*, 4:231.

Latour, B. and Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts*. New York, NY: Sage.

Latour, B. and Woolgar, S. (2013). *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press.

Lawrence, T., Suddaby, R., and Leca, B. (2011). Institutional work: Refocusing institutional studies of organization. *Journal of Management Inquiry*, 20(1):52–58.

Lawrence, T. B. (1999). Institutional strategy. *Journal of Management*, 25(2):161–187.

Lehmann, F. (1992). Semantic networks. *Computers Mathematic Application*, 23(2-5):1–50.

Leydesdorff, L. and Etzkowitz, H. (1996). Emergence of a triple helix of universityindustrygovernment relations. *Science and Public Policy*, 23(5):279–286.

Liben-Nowell, D. and Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal American Society for Information Science and Technology*, 58(7):1019–1031.

Lin, M.-W. and Bozeman, B. (2006). Researchers industry experience and productivity in universityindustry research centers: A scientific and technical human capital explanation. *The Journal of Technology Transfer*, 31(2):269–290.

Long, J. S. (1978). Productivity and academic position in the scientific career. *American Sociological Review*, 889–908.

Lundqvist, A., Apers, P., Smeulders, A., Huizer, E., and Mandersloot, P. (2012). ICT Road Map. http://www.nwo.nl/binaries/content/documents/nwo/algemeen/documentation/application/ew/roadmap-ict-2012.

Lusher, D., Koskinen, J., and Robins, G. (2012). *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge, England: Cambridge University Press.

Martin, J. L. (1999). A general permutation-based QAP analysis approach for dyadic data from multiple groups. *Connections*, 22(2):50–60.

Martin, J. L. (2003). What is field theory? *American Journal of Sociology*, 109(1):1–49.

McEvily, B., Soda, G., and Tortoriello, M. (2014). More formally: Rediscovering the missing link between formal organization and informal social structure. *he Academy of Management Annals*, 8(1):299–345.

McFadyen, M. A., Semadeni, M., and Cannella Jr, A. A. (2009). Value of strong ties to disconnected others: Examining knowledge creation in biomedicine. *Organization Science*, 20(3):552–564.

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444.

Mehra, A., Dixon, A. L., Brass, D. J., and Robertson, B. (2006). The social network ties of group leaders: Implications for group performance and leader reputation. *Organization Science*, 17(1):64–79.

Mehra, A., Kilduff, M., and Brass, D. J. (2001). The social networks of high and low self-monitors: Implications for workplace performance. *Administrative Science Quarterly*, 46(1):121–146.

Melin, G. (2000). Pragmatism and self- organization: research collaboration on the individual level. *Research Policy*, 29(1):31–40.

Melin, G. and Persson, O. (1996). Studying research collaboration using co-authorships. *Scientometrics*, 36(3):363–377.

Merton, R. K. (1968). The matthew effect in science. *Science*, 159(3810):56–63.

Meyer, J. W. and Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 340–363.

Meyer, M. and Rowan, B. (1991). Institutionalised organisations: Formal structures, myth and ceremony. In Powell, W & DiMaggio, P., editor, *The New Institutionalism in Organisational Analysis*. Chicago, IL: University of Chicago Press.

Microsoft Academic Search (2014). http://academic.research.microsoft.com/.

Mika, P., Elfring, T., and Groenewegen, P. L. M. (2006). Application of semantic technology for social network analysis in the sciences. *Scientometrics*, 68(1):3–27.

Moed, H. F. and Visser, M. S. (2007). Developing bibliometric indicators of research performance in computer science: an exploratory study. Technical Report 2007-01, Leiden, The Netherlands: CWTS.

Monge, P. R. and Contractor, N. S. (2003). *Theories of communication networks*. New York, NY: Oxford University Press.

Monge, P. R. and Eisenberg, E. M. (1987). *Emergent communication networks*. New York, NY: Sage.

Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, 69(2):213–238.

Moore, C., Ghoshal, G., and Newman, M. E. J. (2006). Exact solutions for models of evolving networks with addition and deletion of nodes. *Physical Review E*, 74:036121.

Moser, C., Birkholz, J., Deichmann, D., Groenewegen, P., and Wang, S. (2013a). The emergence of successful ideas from social and content networks. In *Proc. of XXXIII Sunbelt Social Networks Conference of the International Network for Social Network Analysis*. University of Hamburg, Hamburg, Germany.

Moser, C., Birkholz, J. M., Deichmann, D., Hellsten, I., and Wang, S. (2013b). Exploring ideation: Knowledge development in science through the lens of semantic and social networks. In *System Sciences (HICSS), 2013 46th Hawaii International Conference on*, 235–243. New York, NY: IEEE.

Müller, A. and Stoyan, D. (2002). *Comparison methods for stochastic models and risks*, volume 389. Hoboken, NJ: Wiley.

NARCIS (2014). National academic research and collaborations information system. http://www.narcis.nl.

Nationale Informaticakamer (2010). Assessment of research quality in computer science, 2002–2008. http://media.leidenuniv.nl/legacy/computer-science-def.pdf.

Network Institute (2014). http://www.networkinstitute.org.

Newell, A. and Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, 19(3):113–126.

Newman, M. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Science*, 101(suppl 1):5200–5205.

Newman, M. E. and Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E*, 68(3):036122.

Newman, M. E. J. (2001a). Scientific collaboration networks I. Network construction and fundamental results. *Physical Review E*, 64:016131.

Newman, M. E. J. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64:016132.

North, D. C. (1991). Institutions. *Journal of Economic Perspectives*, 5(1):97–112.

Obstfeld, D. (2005). Social networks, the tertius iungens orientation, and involvement in innovation. *Administrative Science Quarterly*, 50(1):100–130.

Obstfeld, D., Borgatti, S. P., and Davis, J. (2014). Brokerage as a process: Decoupling third party action from social network structure. *Research in the Sociology of Organizations*, 40:135–159.

Oliver, C. (1991). Strategic responses to institutional processes. *Academy of Management Review*, 16(1):145–179.

Opsahl, T. and Panzarasa, P. (2009). Clustering in weighted networks. *Social Networks*, 31(2):155–163.

Owen-Smith, J., Riccaboni, M., Pammolli, F., and Powell, W. W. (2002). A comparison of us and european university-industry relations in the life sciences. *Management Science*, 48(1):24–43.

Palazzolo, E. T., Serb, D. A., She, Y., Su, C., and Contractor, N. S. (2006). Coevolution of communication and knowledge networks in transactive memory systems: Using computational models for theoretical development. *Communication Theory*, 16(2):223–250.

Pfeffer, J. and Salancik, G. R. (2003). *The external control of organizations: A resource dependence perspective*. Palo Alto, CA: Stanford University Press.

Podolny, J. M. (2001). Networks as the pipes and prisms of the market1. *American Journal of Sociology*, 107(1):33–60.

Polanyi, M. (1967). *The tacit dimension*. London, England: Routledge & Kegan Paul.

Ponomariov, B. L. (2007). Effects of university characteristics on scientists' interactions with the private sector: an exploratory assessment. *The Journal of Technology Transfer*, 33(5):485–503.

Ponomariov, B. L. and Boardman, P. C. (2010). Influencing scientists' collaboration and productivity patterns through new institutions: University research centers and scientific and technical human capital. *Research Policy*, 39(5):613–624.

Poole, M. S. and Contractor, N. (2011). Conceptualizing the multiteam system as an ecosystem of networked groups. In Zaccaro, S. J., Marks, M. A., and DeChurch, L., editors, *Multiteam systems: An organizational form for dynamic and complex environments*, 193–224. London, England: Rotuledge.

Popescul, A. and Ungar, L. (2003). Statistical relational learning for link prediction. In *Proceedings of the Conference on Artificial Intelligence*, 81–90. New York, NY: ACM.

Popping, R. (2003). Knowledge Graphs and Network Text Analysis. *Social Science Information*, 42(1):91–106.

Powell, W. W. (1990). Neither market nor hierarchy: Network forms of organization. *Research in Organizational Behavior*, 12:295–336.

Price, D. d. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27(5):292–306.

Price, D. J. (1965). Networks of scientific papers. *Science*, 149:510.

Rabe-Hesketh, S. and Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*. College Station, Texas: STATA press.

Rabiner, L. R. (1990). A tutorial on hidden markov models and selected applications in speech recognition. In *Readings in Speech Recognition*, 267–296. Burlington, MA: Morgan Kaufmann.

Radicchi, F., Fortunato, S., and Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences*, 105(45):17268–17272.

Rathenau Institute (2014a). FES-programma's voor kennis en innovatie. http://www.rathenau.nl/nc/web-specials/de-nederlandse-wetenschap/financiering/overheid/fes-programmas-voor-kennis-en-innovatie.html.

Rathenau Institute (2014b). Universiteiten en UMC's. http://www.rathenau.nl/nc/web-specials/de-nederlandse-wetenschap/organisaties/universiteiten-en-umcs.html.

Reagans, R. and McEvily, B. (2003). Network structure and knowledge transfer: The effects of cohesion and range. *Administrative Science Quarterly*, 48(2):240–267.

Ripley, R. M., Snijders, T. A., and Preciado, P. (2011). Manual for SIENA version 4.0. *Oxford, England: University of Oxford*.

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2):173–191.

Rodriguez, M. a. and Pepe, A. (2008). On the relationship between the structural and socioacademic communities of a coauthorship network. *Journal of Informetrics*, 2(3):195–201.

Salancik, G. R. (1995). Wanted: A good network theory of organization. *Administrative Science Quarterly*, 40(2):345–349.

Sauer, R. D. (1988). Estimates of the returns to quality and coauthorship in economic academia. *Journal of Political Economy*, 96(4):855–866.

Scott, W. R. (1995). *Institutions and Organizations*. New York, NY: Sage.

Sewell Jr, W. H. (1992). A theory of structure: Duality, agency, and transformation. *American Journal of Sociology*, 98(1):1–29.

Shah, P. P. (1998). Who are employees' social referents? Using a network perspective to determine referent others. *Academy of Management Journal*, 41(3):249–268.

Shah, P. P. (2000). Network destruction: The structural implications of downsizing. *Academy of Management Journal*, 43(1):101–112.

Shrum, W., Genuth, J., and Chompalov, I. (2007). *Structures of Scientific Collaboration*. Cambridge, MA: MIT Press.

Shumate, M. and Contractor, N. (2013). Emergence of multidimensional social networks. In Putnam, L. L. and Mumby, D. K., editors, *The SAGE handbook of organizational communication*, 449–474. New York, NY: Sage.

SIKS (2014). http://www.siks.nl/.

Simmel, G. (1955). The web of group affiliations. *Conflict and the web of group affiliations*, 125–95.

Simmel, G. and Wolff, K. H. (1950). *The sociology of Georg Simmel*. New York, NY: The Free Press.

Snijders, T. (2005). Models for longitudinal network data. In *Models and methods in social network analysis*, chapter 11, 215–247. Cambridge, England: Cambridge University Press.

Snijders, T. A. and Baerveldt, C. (2003). A multilevel network study of the effects of delinquent behavior on friendship evolution. *Journal of Mathematical Sociology*, 27(2-3):123–151.

Snijders, T. A., Van de Bunt, G. G., and Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1):44–60.

Stokols, D., Hall, K. L., Taylor, B. K., and Moser, R. P. (2008a). The science of team science: overview of the field and introduction to the supplement. *American Journal of Preventive Medicine*, 35(2):S77–S89.

Stokols, D., Misra, S., Moser, R., Hall, K., and Taylor, B. (2008b). The ecology of team science: understanding contextual influences on transdisciplinary collaboration. *American Journal Preventive Medicine*, 35(2S):S96–S115.

Taskar, B., Wong, M.-F., Abbeel, P., and Koller, D. (2003). Link prediction in relational data. In *Proceedings of Neural Information Processing Systems*, 659–666. Cambridge, MA: MIT Press.

Taylor, M. and Doerfel, M. L. (2003). Building interorganizational relationships that build nations. *Human Communication Research*, 29(2):153–181.

Tedre, M. (2006). The development of computer science: A sociocultural perspective. In *Proceedings of the 6th Baltic Sea conference on Computing education research: Koli Calling 2006*, 21–24. New York, NY: ACM.

Teece, D. J., Pisano, G., and Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, 18(7):509–533.

Thomas, G. (2010). *How to do your case study: A guide for students and researchers*. New York, NY: Sage.

Tichy, N. and Fombrun, C. (1979). Network analysis in organizational settings. *Human Relations*, 32(11):923–965.

Tortoriello, M. and Krackhardt, D. (2010). Activating cross-boundary knowledge: the role of simmelian ties in the generation of innovations. *Academy of Management Journal*, 53(1):167–181.

Toulmin, S. E. (1972). *Human understanding*. Princeton, NJ: Princeton University Press.

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, 42:35–67.

Uzzi, B. (2008). A social network's changing statistical properties and the quality of human innovation. *Journal of Physics A: Mathematical and Theoretical*, 41(22):224023.

Van den Besselaar, P. and Leydesdorff, L. (1996). Mapping change in scientific specialties: A scientometric reconstruction of the development of artificial intelligence. *Journal of the American Society for Information Science*, 47(6):415–436.

Van den Bulte, C. and Moenaert, R. K. (1998). The effects of R&D team co-address on communication patterns among R&D, marketing, and manufacturing. *Management Science*, 44(11, Part 2):S1–S18.

van den Toren, J. P., Hessels, L., Eveleens, C., and van der Meulen, B. (2012). Cordinatie in de topsectoren. de geplande tki's en hun uitdagingen. Technical report, Leiden, The Netherlands: Rathenau Institute.

van Steen, M. (2010). *An Introduction to Graph Theory and Complex Networks*. Amazon.

Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications*, volume 8 of *Structural Analysis in the Social Sciences*. Cambridge, England: Cambridge University Press.

Watts, D. J. (1999). Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, 105:493–527.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small- world' networks. *Nature*, 393(49):440–442.

Wegner, P. (1976). Research paradigms in computer science. *Proceedings of the 2nd international Conference on Software Engineering*, 322–330.

Weinstock, M. (1971). Citation indexes. In Kent, A., editor, *Encyclopedia of Library and Information Science*, 16–41. New York, NY: Marcel Dekker.

Weiss, R. S. (1995). *Learning from Strangers: The Art and Method of Qualitative Interview Studies*. New York, NY: Simon and Schuster.

Whitbred, R., Fonti, F., Steglich, C., and Contractor, N. (2011). From microactions to macrostructure and back: A structurational approach to the evolution of organizational networks. *Human Communication Research*, 37(3):404–433.

Whitley, R. (2000). *The intellectual and social organization of the sciences*. Cambridge, England: Oxford University Press.

Williams, D., Contractor, N., Poole, M. S., Srivastava, J., and Cai, D. (2011). The virtual worlds exploratorium: using large-scale data and computational techniques for communication research. *Communication Methods and Measures*, 5(2):163–180.

Wooten, M. and Hoffman, A. J. (2008). Organizational fields: Past, present and future. *The Sage Handbook of Organizational Institutionalism*, 130–147.

Wouters, P. F. (1999). *The citation culture*. PhD thesis, Universiteit van Amsterdam.

Wuchty, S., Jones, B. F., and Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–1039.

Yahoo! PlaceFinder (2012). http://developer.yahoo.com/geo/placefinder/.

Zucker, L. G. (1977). The role of institutionalization in cultural persistence. *American Sociological Review*, 726–743.

# Appendix

## Dutch Computer Science Departmental Units Chapter 2

Table A.1: Dutch Computer Science Department Units

| University (in no particular order) | Organization within the university | Groups affiliated with this case | Website |
|---|---|---|---|
| Technische Universiteit Delft | Departments: Intelligent Systems, & Software and Computer Technology | Algorithmics Computer Engineering, Computer Graphics and Visualization, Cyber Security, Embedded Software, Interactive Intelligence, Multimedia Signal Processing, Network Architectures and Systems, Parallel and Distributed Systems, Pattern Recognition and Bioinformatics, Software Engineering, Web Information Systems | www.ewi.tudelft.nl/over-de-faculteit/computer-science-engineering/ |
| Vrije Universiteit Amsterdam | Department of Computer Sciences | Artificial Intelligence, Bioinformatics, Business Web & Media, Computer Systems, Information Management & Software Engineering, Theoretical Computer Science | www.cs.vu.nl |
| Technische Universiteit Eindhoven | Department of Mathematics and Computer Science | Algorithms and Visualization, Information Systems, Model Driven Software Engineering, Security and Embedded Networked Systems | www.tue.nl/en/university/departments/mathematics-and-computer-science/ |
| Universiteit van Twente | Discipline: Informatica | Computer Architecture for Embedded Systems; Design and Analysis of Communication Systems; Databases; Formal Methods and Tools; Human Media Interaction; Pervasive Systems; Services, Cybersecurity & Safety | www.utwente.nl/onderwijs/ewi/ |
| Universiteit Leiden | Leiden Institute of Advanced Computer Science | Algorithms and Software Technology, Computer Systems and Imagery & Media | www.liacs.nl/ |

*Continued on next page*

Table A.1 – *Continued from previous page*

| University (in no particular order) | Organization within the university | Groups affiliated with this case | Website |
|---|---|---|---|
| Radboud Nijmegen Universiteit | Institute for Computing and Information Sciences | Model Based System Development, Digital Security, Intelligent Systems | www.ru.nl/icis/ |
| Universiteit van Amsterdam | Institute for Informatics * | Computer Systems Architecture, Intelligent Autonomous Systems, Information and Language Processing Systems, Intelligent Sensory Information Systems, Computational Science, Theory of Computer Science, Federated Collaborative Networks, System and Network Engineering, Lecturers group** | ivi.uva.nl/ |
| Universiteit Utrecht | Department of Information and Computing Sciences | Games and Virtual Worlds, Multimedia and Geometry, Intelligent Systems, Decision Support Systems, Algorithmic Data Analysis, Algorithmic System, Software Technology, Organisation and Information | www.cs.uu.nl/ |
| Rijksuniversiteit Groningen | Institutes: Johann Bernoulli Institute of Mathematics and Computing Science, & Institute of Artificial Intelligence and Cognitive Engineering (ALICE) | Bioinformatics, Distributed Systems, Fundamental Computing, Intelligent Systems, Scientific Visualization and Computer Graphics, Software Engineering, Autonomous Perceptive Systems, Cognitive Modeling, Multi-agent Systems, Sensory Cognition | www.rug.nl/research/fmns/ |

* not to be confused with Centrum Wiskunde & Informatica (CWI), which is not an academic department within a university but research institute, both the UvA & CWI are physically located at Amsterdam Science Park and cooperate both formally and informally, thus some scientists within this study may have dual affiliations.

** Special distinction, a group that just has lecturing responsibilities, no research.

## Means, standard deviations of variables, and correlation matrix Chapter 3

See the next page.

Table A.2: Means, standard deviations of variables, and correlation matrix (SF. stands for Subfield, M – Mean)

| **Variable** | $N$ | M | SD | min | max | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Performance Success (Citation scores) | 193 | 2.73 | 2.22 | 0 | 7.79 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Constraint | 194 | .51 | .25 | .14 | 1.13 | -0.55* | | | | | | | | | | | | | | | | | | | | | | | | | |
| Target list | 194 | .29 | .45 | 0 | 1 | -0.08 | .09 | | | | | | | | | | | | | | | | | | | | | | | | |
| Professional tenure system | 194 | .10 | .30 | 0 | 1 | -0.02 | .08 | -0.02 | | | | | | | | | | | | | | | | | | | | | | | |
| Position | 194 | 2.52 | 1.47 | 0 | 5 | .54* | -0.41* | -0.22* | .13 | | | | | | | | | | | | | | | | | | | | | | |
| Gender | 194 | .14 | .35 | 0 | 1 | .03 | .01 | .14 | .02 | -0.07 | | | | | | | | | | | | | | | | | | | | | |
| Dutch nationality | 194 | .70 | .46 | 0 | 1 | .06 | .07 | -0.05 | .03 | .16* | -0.22* | | | | | | | | | | | | | | | | | | | | |
| SF. applied mathematics | 194 | .03 | .17 | 0 | 1 | -0.15* | .21* | -0.05 | -0.06 | .02 | .01 | -0.01 | | | | | | | | | | | | | | | | | | | |
| SF. artificial intelligence | 194 | .26 | .44 | 0 | 1 | .00 | -0.04 | .11 | .00 | -0.14 | .10 | -0.09 | -0.11 | | | | | | | | | | | | | | | | | | |
| SF. bioinformatics | 194 | .03 | .16 | 0 | 1 | .01 | .01 | -0.10 | .06 | .01 | -0.07 | .11 | -0.03 | -0.10 | | | | | | | | | | | | | | | | | |
| SF. tradit. computer studies | 194 | .20 | .40 | 0 | 1 | .08 | .01 | -0.14* | .06 | .05 | -0.09 | .13 | -0.09 | -0.29* | -0.08 | | | | | | | | | | | | | | | | |
| SF. databases management | 194 | .04 | .19 | 0 | 1 | .02 | .07 | .06 | -0.06 | .03 | .00 | -0.05 | -0.03 | -0.12 | -0.03 | -0.10 | | | | | | | | | | | | | | | |
| SF. image sound | 194 | .04 | .19 | 0 | 1 | .00 | -0.07 | .00 | .12 | .06 | .00 | -0.05 | -0.03 | -0.12 | -0.03 | -0.10 | -0.04 | | | | | | | | | | | | | | |
| SF. web development design | 194 | .03 | .16 | 0 | 1 | -0.04 | -0.12 | .04 | -0.05 | -0.01 | -0.07 | -0.03 | -0.03 | -0.10 | -0.03 | -0.08 | -0.03 | -0.03 | | | | | | | | | | | | | |
| SF. networks | 194 | .10 | .30 | 0 | 1 | -0.02 | .03 | .08 | .00 | -0.05 | .01 | .00 | -0.06 | -0.20* | -0.06 | -0.17* | -0.07 | -0.07 | -0.06 | | | | | | | | | | | | |
| SF. operating systems | 194 | .02 | .14 | 0 | 1 | -0.05 | .03 | -0.01 | -0.05 | -0.05 | -0.06 | -0.06 | -0.03 | -0.09 | -0.02 | -0.07 | -0.03 | -0.03 | -0.02 | -0.05 | | | | | | | | | | | |
| SF. simulations | 194 | .11 | .32 | 0 | 1 | -0.04 | -0.02 | -0.12 | -0.06 | -0.03 | .09 | -0.01 | -0.06 | -0.21* | -0.06 | -0.18* | -0.07 | -0.07 | -0.06 | -0.12 | -0.05 | | | | | | | | | | |
| SF. software systems | 194 | .13 | .34 | 0 | 1 | .05 | .00 | .08 | -0.03 | .12 | -0.11 | .03 | -0.07 | -0.23* | -0.06 | -0.19* | -0.08 | -0.08 | -0.06 | -0.13 | -0.06 | -0.14 | | | | | | | | | |
| SF. other | 194 | .02 | .12 | 0 | 1 | .01 | -0.11 | .01 | .10 | .10 | .19* | -0.01 | -0.02 | -0.07 | -0.02 | -0.06 | -0.02 | -0.02 | -0.02 | -0.04 | -0.02 | -0.04 | -0.05 | | | | | | | | |
| Affil. with research institute | 194 | .37 | .48 | 0 | 1 | -0.14* | .07 | .08 | -0.11 | -0.09 | -0.06 | .15* | -0.01 | -0.02 | .15* | -0.02 | -0.03 | -0.09 | .08 | .02 | -0.11 | .07 | .02 | -0.10 | | | | | | | |
| Experience outside academia | 194 | .52 | .50 | 0 | 1 | .07 | -0.09 | .07 | .04 | .07 | .00 | -0.06 | -0.01 | -0.05 | .16* | -0.04 | .02 | -0.09 | .03 | -0.04 | .07 | -0.01 | .05 | .12 | .05 | | | | | | |
| # solo pubs (ln) | 194 | .18 | .47 | 0 | 2.77 | .31* | -0.18* | -0.02 | .01 | .30* | .01 | .05 | -0.02 | .11 | -0.06 | .16* | -0.07 | -0.07 | -0.01 | -0.06 | -0.06 | -0.11 | .01 | -0.05 | -0.11 | .03 | | | | | |
| # pubs before 2006 (ln) | 194 | 1.01 | 1.37 | 0 | 4.91 | .60* | -0.37* | -0.11 | -0.02 | .63* | -0.10 | .09 | -0.11 | -0.08 | -0.08 | .03 | .12 | .07 | .07 | -0.04 | -0.04 | -0.02 | .07 | .08 | -0.18* | .06 | .38* | | | | |
| # pubs after 2006 (ln) | 194 | 2.08 | 1.03 | .69 | 4.71 | .66* | -0.58* | -0.21* | -0.11 | .55* | .00 | -0.01 | -0.13 | .09 | -0.05 | .06 | .01 | -0.06 | .14* | -0.04 | -0.05 | -0.08 | -0.02 | .01 | -0.12 | .05 | .44* | .60* | | | |
| # co-authors before 2006 (ln) | 194 | 1.04 | 1.39 | 0 | 4.64 | .53* | -0.38* | -0.10 | -0.01 | .60* | -0.13 | .06 | -0.11 | -0.10 | -0.06 | -0.01 | .11 | .13 | .05 | .05 | -0.04 | -0.05 | .07 | .07 | -0.18* | .08 | .25* | .93* | .53* | | |
| # co-authors after 2006 (ln) | 194 | 2.47 | .95 | .69 | 5.30 | .60* | -0.81* | -0.17* | -0.16* | .46* | .02 | -0.10 | -0.19* | .11 | -0.03 | -0.04 | -0.02 | .03 | .17* | -0.03 | -0.04 | .00 | -0.03 | .00 | -0.05 | .06 | .30* | .52* | .83* | .52* | |
| Citations before 2006 (ln) | 193 | 1.78 | 2.23 | 0 | 7.76 | .84* | -0.38* | -0.10 | .01 | .58* | -0.02 | .16* | -0.13 | -0.07 | .07 | .11 | .06 | .03 | -0.07 | -0.01 | -0.06 | -0.03 | .04 | .03 | -0.15* | .02 | .28* | .72* | .51* | .64* | .44* |

# SIKS Dissertations

## 2009

**2009-01** Rasa Jurgelenaite (RUN), *Symmetric Causal Independence Models.*
**2009-02** Willem Robert van Hage (VU), *Evaluating Ontology-Alignment Techniques.*
**2009-03** Hans Stol (UvT), *A Framework for Evidence-based Policy Making Using IT.*
**2009-04** Josephine Nabukenya (RUN), *Improving the Quality of Organisational Policy Making using Collaboration Engineering.*
**2009-05** Sietse Overbeek (RUN), *Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality.*
**2009-06** Muhammad Subianto (UU), *Understanding Classification.*
**2009-07** Ronald Poppe (UT), *Discriminative Vision-Based Recovery and Recognition of Human Motion.*
**2009-08** Volker Nannen (VU), *Evolutionary Agent-Based Policy Analysis in Dynamic Environments.*
**2009-09** Benjamin Kanagwa (RUN), *Design, Discovery and Construction of Service-oriented Systems.*
**2009-10** Jan Wielemaker (UVA), *Logic programming for knowledge-intensive interactive applications.*
**2009-11** Alexander Boer (UVA), *Legal Theory, Sources of Law & the Semantic Web.*
**2009-12** Peter Massuthe (TUE, Humboldt-Universitaet zu Berlin), *Operating Guidelines for Services.*
**2009-13** Steven de Jong (UM), *Fairness in Multi-Agent Systems.*
**2009-14** Maksym Korotkiy (VU), *From ontology-enabled services to service-enabled ontologies (making ontologies work in e-science with ONTO-SOA).*
**2009-15** Rinke Hoekstra (UVA), *Ontology Representation - Design Patterns and Ontologies that Make Sense.*
**2009-16** Fritz Reul (UvT), *New Architectures in Computer Chess.*

**2009-17** Laurens van der Maaten (UvT), *Feature Extraction from Visual Data.*
**2009-18** Fabian Groffen (CWI), *Armada, An Evolving Database System.*
**2009-19** Valentin Robu (CWI), *Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets.*
**2009-20** Bob van der Vecht (UU), *Adjustable Autonomy: Controling Influences on Decision Making.*
**2009-21** Stijn Vanderlooy (UM), *Ranking and Reliable Classification.*
**2009-22** Pavel Serdyukov (UT), *Search For Expertise: Going beyond direct evidence.*
**2009-23** Peter Hofgesang (VU), *Modelling Web Usage in a Changing Environment.*
**2009-24** Annerieke Heuvelink (VUA), *Cognitive Models for Training Simulations.*
**2009-25** Alex van Ballegooij (CWI), *"RAM: Array Database Management through Relational Mapping".*
**2009-26** Fernando Koch (UU), *An Agent-Based Model for the Development of Intelligent Mobile Services.*
**2009-27** Christian Glahn (OU), *Contextual Support of social Engagement and Reflection on the Web.*
**2009-28** Sander Evers (UT), *Sensor Data Management with Probabilistic Models.*
**2009-29** Stanislav Pokraev (UT), *Model-Driven Semantic Integration of Service-Oriented Applications.*
**2009-30** Marcin Zukowski (CWI), *Balancing vectorized query execution with bandwidth-optimized storage.*
**2009-31** Sofiya Katrenko (UVA), *A Closer Look at Learning Relations from Text.*
**2009-32** Rik Farenhorst (VU) and Remco de Boer (VU), *Architectural Knowledge Management: Supporting Architects and Auditors.*
**2009-33** Khiet Truong (UT), *How Does Real Affect Affect Affect Recognition In Speech?.*
**2009-34** Inge van de Weerd (UU), *Advancing in Software Product Management: An Incremental Method Engineering Approach.*
**2009-35** Wouter Koelewijn (UL), *Privacy en Politiegegevens; Over geautomatiseerde normatieve informatie-uitwisseling.*
**2009-36** Marco Kalz (OUN), *Placement Support for Learners in Learning Networks.*
**2009-37** Hendrik Drachsler (OUN), *Navigation Support for Learners in Informal Learning Networks.*
**2009-38** Riina Vuorikari (OU), *Tags and self-organisation: a metadata ecology for learning resources in a multilingual context.*
**2009-39** Christian Stahl (TUE, Humboldt-Universitaet zu Berlin), *Service Substitution – A Behavioral Approach Based on Petri Nets.*
**2009-40** Stephan Raaijmakers (UvT), *Multinomial Language Learning: Investigations into the Geometry of Language.*
**2009-41** Igor Berezhnyy (UvT), *Digital Analysis of Paintings.*
**2009-42** Toine Bogers (UvT), *Recommender Systems for Social Bookmarking.*
**2009-43** Virginia Nunes Leal Franqueira (UT), *Finding Multi-step Attacks in Com-*

*puter Networks using Heuristic Search and Mobile Ambients.*
**2009-44** Roberto Santana Tapia (UT), *Assessing Business-IT Alignment in Networked Organizations.*
**2009-45** Jilles Vreeken (UU), *Making Pattern Mining Useful.*
**2009-46** Loredana Afanasiev (UvA), *Querying XML: Benchmarks and Recursion.*

## 2010

**2010-01** Matthijs van Leeuwen (UU), *Patterns that Matter.*
**2010-02** Ingo Wassink (UT), *Work flows in Life Science.*
**2010-03** Joost Geurts (CWI), *A Document Engineering Model and Processing Framework for Multimedia documents.*
**2010-04** Olga Kulyk (UT), *Do You Know What I Know? Situational Awareness of Co-located Teams in Multidisplay Environments.*
**2010-05** Claudia Hauff (UT), *Predicting the Effectiveness of Queries and Retrieval Systems.*
**2010-06** Sander Bakkes (UvT), *Rapid Adaptation of Video Game AI.*
**2010-07** Wim Fikkert (UT), *Gesture interaction at a Distance.*
**2010-08** Krzysztof Siewicz (UL), *Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments.*
**2010-09** Hugo Kielman (UL), *A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging.*
**2010-10** Rebecca Ong (UL), *Mobile Communication and Protection of Children.*
**2010-11** Adriaan Ter Mors (TUD), *The world according to MARP: Multi-Agent Route Planning.*
**2010-12** Susan van den Braak (UU), *Sensemaking software for crime analysis.*
**2010-13** Gianluigi Folino (RUN), *High Performance Data Mining using Bio-inspired techniques.*
**2010-14** Sander van Splunter (VU), *Automated Web Service Reconfiguration.*
**2010-15** Lianne Bodenstaff (UT), *Managing Dependency Relations in Inter-Organizational Models.*
**2010-16** Sicco Verwer (TUD), *Efficient Identification of Timed Automata, theory and practice.*
**2010-17** Spyros Kotoulas (VU), *Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications.*
**2010-18** Charlotte Gerritsen (VU), *Caught in the Act: Investigating Crime by Agent-Based Simulation.*
**2010-19** Henriette Cramer (UvA), *People's Responses to Autonomous and Adaptive Systems.*
**2010-20** Ivo Swartjes (UT), *Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative.*

**2010-21** Harold van Heerde (UT), *Privacy-aware data management by means of data degradation.*
**2010-22** Michiel Hildebrand (CWI), *End-user Support for Access to Heterogeneous Linked Data.*
**2010-23** Bas Steunebrink (UU), *The Logical Structure of Emotions.*
**2010-24** Dmytro Tykhonov, *Designing Generic and Efficient Negotiation Strategies.*
**2010-25** Zulfiqar Ali Memon (VU), *Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective.*
**2010-26** Ying Zhang (CWI), *XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines.*
**2010-27** Marten Voulon (UL), *Automatisch contracteren.*
**2010-28** Arne Koopman (UU), *Characteristic Relational Patterns.*
**2010-29** Stratos Idreos(CWI), *Database Cracking: Towards Auto-tuning Database Kernels.*
**2010-30** Marieke van Erp (UvT), *Accessing Natural History - Discoveries in data cleaning, structuring, and retrieval.*
**2010-31** Victor de Boer (UVA), *Ontology Enrichment from Heterogeneous Sources on the Web.*
**2010-32** Marcel Hiel (UvT), *An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems.*
**2010-33** Robin Aly (UT), *Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval.*
**2010-34** Teduh Dirgahayu (UT), *Interaction Design in Service Compositions.*
**2010-35** Dolf Trieschnigg (UT), *Proof of Concept: Concept-based Biomedical Information Retrieval.*
**2010-36** Jose Janssen (OU), *Paving the Way for Lifelong Learning; Facilitating competence development through a learning path specification.*
**2010-37** Niels Lohmann (TUE), *Correctness of services and their composition.*
**2010-38** Dirk Fahland (TUE), *From Scenarios to components.*
**2010-39** Ghazanfar Farooq Siddiqui (VU), *Integrative modeling of emotions in virtual agents.*
**2010-40** Mark van Assem (VU), *Converting and Integrating Vocabularies for the Semantic Web.*
**2010-41** Guillaume Chaslot (UM), *Monte-Carlo Tree Search.*
**2010-42** Sybren de Kinderen (VU), *Needs-driven service bundling in a multi-supplier setting - the computational e3-service approach.*
**2010-43** Peter van Kranenburg (UU), *A Computational Approach to Content-Based Retrieval of Folk Song Melodies.*
**2010-44** Pieter Bellekens (TUE), *An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain.*
**2010-45** Vasilios Andrikopoulos (UvT), *A theory and model for the evolution of software services.*
**2010-46** Vincent Pijpers (VU), *e3alignment: Exploring Inter-Organizational Business-*

*ICT Alignment.*
**2010-47** Chen Li (UT), *Mining Process Model Variants: Challenges, Techniques, Examples.*
**2010-48** Withdrawn, .
**2010-49** Jahn-Takeshi Saito (UM), *Solving difficult game positions.*
**2010-50** Bouke Huurnink (UVA), *Search in Audiovisual Broadcast Archives.*
**2010-51** Alia Khairia Amin (CWI), *Understanding and supporting information seeking tasks in multiple sources.*
**2010-52** Peter-Paul van Maanen (VU), *Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention.*
**2010-53** Edgar Meij (UVA), *Combining Concepts and Language Models for Information Access.*

## 2011

**2011-01** Botond Cseke (RUN), *Variational Algorithms for Bayesian Inference in Latent Gaussian Models.*
**2011-02** Nick Tinnemeier(UU), *Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language.*
**2011-03** Jan Martijn van der Werf (TUE), *Compositional Design and Verification of Component-Based Information Systems.*
**2011-04** Hado van Hasselt (UU), *Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference.*
**2011-05** Base van der Raadt (VU), *Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline..*
**2011-06** Yiwen Wang (TUE), *Semantically-Enhanced Recommendations in Cultural Heritage.*
**2011-07** Yujia Cao (UT), *Multimodal Information Presentation for High Load Human Computer Interaction.*
**2011-08** Nieske Vergunst (UU), *BDI-based Generation of Robust Task-Oriented Dialogues.*
**2011-09** Tim de Jong (OU), *Contextualised Mobile Media for Learning.*
**2011-10** Bart Bogaert (UvT), *Cloud Content Contention.*
**2011-11** Dhaval Vyas (UT), *Designing for Awareness: An Experience-focused HCI Perspective.*
**2011-12** Carmen Bratosin (TUE), *Grid Architecture for Distributed Process Mining.*
**2011-13** Xiaoyu Mao (UvT), *Airport under Control. Multiagent Scheduling for Airport Ground Handling.*
**2011-14** Milan Lovric (EUR), *Behavioral Finance and Agent-Based Artificial Markets.*
**2011-15** Marijn Koolen (UvA), *The Meaning of Structure: the Value of Link Evi-*

*dence for Information Retrieval.*
**2011-16** Maarten Schadd (UM), *Selective Search in Games of Different Complexity.*
**2011-17** Jiyin He (UVA), *Exploring Topic Structure: Coherence, Diversity and Relatedness.*
**2011-18** Mark Ponsen (UM), *Strategic Decision-Making in complex games.*
**2011-19** Ellen Rusman (OU), *The Mind ' s Eye on Personal Profiles.*
**2011-20** Qing Gu (VU), *Guiding service-oriented software engineering - A view-based approach.*
**2011-21** Linda Terlouw (TUD), *Modularization and Specification of Service-Oriented Systems.*
**2011-22** Junte Zhang (UVA), *System Evaluation of Archival Description and Access.*
**2011-23** Wouter Weerkamp (UVA), *Finding People and their Utterances in Social Media.*
**2011-24** Herwin van Welbergen (UT), *Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior.*
**2011-25** Syed Waqar ul Qounain Jaffry (VU)), *Analysis and Validation of Models for Trust Dynamics.*
**2011-26** Matthijs Aart Pontier (VU), *Virtual Agents for Human Communication - Emotion Regulation and Involvement-Distance Trade-Offs in Embodied Conversational Agents and Robots.*
**2011-27** Aniel Bhulai (VU), *Dynamic website optimization through autonomous management of design patterns.*
**2011-28** Rianne Kaptein(UVA), *Effective Focused Retrieval by Exploiting Query Context and Document Structure.*
**2011-29** Faisal Kamiran (TUE), *Discrimination-aware Classification.*
**2011-30** Egon van den Broek (UT), *Affective Signal Processing (ASP): Unraveling the mystery of emotions.*
**2011-31** Ludo Waltman (EUR), *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality.*
**2011-32** Nees-Jan van Eck (EUR), *Methodological Advances in Bibliometric Mapping of Science.*
**2011-33** Tom van der Weide (UU), *Arguing to Motivate Decisions.*
**2011-34** Paolo Turrini (UU), *Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations.*
**2011-35** Maaike Harbers (UU), *Explaining Agent Behavior in Virtual Training.*
**2011-36** Erik van der Spek (UU), *Experiments in serious game design: a cognitive approach.*
**2011-37** Adriana Burlutiu (RUN), *Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference.*
**2011-38** Nyree Lemmens (UM), *Bee-inspired Distributed Optimization.*
**2011-39** Joost Westra (UU), *Organizing Adaptation using Agents in Serious Games.*

**2011-40** Viktor Clerc (VU), *Architectural Knowledge Management in Global Software Development.*
**2011-41** Luan Ibraimi (UT), *Cryptographically Enforced Distributed Data Access Control.*
**2011-42** Michal Sindlar (UU), *Explaining Behavior through Mental State Attribution.*
**2011-43** Henk van der Schuur (UU), *Process Improvement through Software Operation Knowledge.*
**2011-44** Boris Reuderink (UT), *Robust Brain-Computer Interfaces.*
**2011-45** Herman Stehouwer (UvT), *Statistical Language Models for Alternative Sequence Selection.*
**2011-46** Beibei Hu (TUD), *Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work.*
**2011-47** Azizi Bin Ab Aziz(VU), *Exploring Computational Models for Intelligent Support of Persons with Depression.*
**2011-48** Mark Ter Maat (UT), *Response Selection and Turn-taking for a Sensitive Artificial Listening Agent.*
**2011-49** Andreea Niculescu (UT), *Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality.*

## 2012

**2012-01** Terry Kakeeto (UvT), *Relationship Marketing for SMEs in Uganda.*
**2012-02** Muhammad Umair(VU), *Adaptivity, emotion, and Rationality in Human and Ambient Agent Models.*
**2012-03** Adam Vanya (VU), *Supporting Architecture Evolution by Mining Software Repositories.*
**2012-04** Jurriaan Souer (UU), *Development of Content Management System-based Web Applications.*
**2012-05** Marijn Plomp (UU), *Maturing Interorganisational Information Systems.*
**2012-06** Wolfgang Reinhardt (OU), *Awareness Support for Knowledge Workers in Research Networks.*
**2012-07** Rianne van Lambalgen (VU), *When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions.*
**2012-08** Gerben de Vries (UVA), *Kernel Methods for Vessel Trajectories.*
**2012-09** Ricardo Neisse (UT), *Trust and Privacy Management Support for Context-Aware Service Platforms.*
**2012-10** David Smits (TUE), *Towards a Generic Distributed Adaptive Hypermedia Environment.*
**2012-11** J.C.B. Rantham Prabhakara (TUE), *Process Mining in the Large: Preprocessing, Discovery, and Diagnostics.*
**2012-12** Kees van der Sluijs (TUE), *Model Driven Design and Data Integration in*

*Semantic Web Information Systems.*
**2012-13** Suleman Shahid (UvT), *Fun and Face: Exploring non-verbal expressions of emotion during playful interactions.*
**2012-14** Evgeny Knutov(TUE), *Generic Adaptation Framework for Unifying Adaptive Web-based Systems.*
**2012-15** Natalie van der Wal (VU), *Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes..*
**2012-16** Fiemke Both (VU), *Helping people by understanding them - Ambient Agents supporting task execution and depression treatment.*
**2012-17** Amal Elgammal (UvT), *Towards a Comprehensive Framework for Business Process Compliance.*
**2012-18** Eltjo Poort (VU), *Improving Solution Architecting Practices.*
**2012-19** Helen Schonenberg (TUE), *What's Next? Operational Support for Business Process Execution.*
**2012-20** Ali Bahramisharif (RUN), *Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing.*
**2012-21** Roberto Cornacchia (TUD), *Querying Sparse Matrices for Information Retrieval.*
**2012-22** Thijs Vis (UvT), *Intelligence, politie en veiligheidsdienst: verenigbare grootheden?.*
**2012-23** Christian Muehl (UT), *Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction.*
**2012-24** Laurens van der Werff (UT), *Evaluation of Noisy Transcripts for Spoken Document Retrieval.*
**2012-25** Silja Eckartz (UT), *Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application.*
**2012-26** Emile de Maat (UVA), *Making Sense of Legal Text.*
**2012-27** Hayrettin Gurkok (UT), *Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games.*
**2012-28** Nancy Pascall (UvT), *Engendering Technology Empowering Women.*
**2012-29** Almer Tigelaar (UT), *Peer-to-Peer Information Retrieval.*
**2012-30** Alina Pommeranz (TUD), *Designing Human-Centered Systems for Reflective Decision Making.*
**2012-31** Emily Bagarukayo (RUN), *A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure.*
**2012-32** Wietske Visser (TUD), *Qualitative multi-criteria preference representation and reasoning.*
**2012-33** Rory Sie (OUN), *Coalitions in Cooperation Networks (COCOON).*
**2012-34** Pavol Jancura (RUN), *Evolutionary analysis in PPI networks and applications.*
**2012-35** Evert Haasdijk (VU), *Never Too Old To Learn – On-line Evolution of Controllers in Swarm- and Modular Robotics.*
**2012-36** Denis Ssebugwawo (RUN), *Analysis and Evaluation of Collaborative Modeling Processes.*

**2012-37** Agnes Nakakawa (RUN), *A Collaboration Process for Enterprise Architecture Creation.*
**2012-38** Selmar Smit (VU), *Parameter Tuning and Scientific Testing in Evolutionary Algorithms.*
**2012-39** Hassan Fatemi (UT), *Risk-aware design of value and coordination networks.*
**2012-40** Agus Gunawan (UvT), *Information Access for SMEs in Indonesia.*
**2012-41** Sebastian Kelle (OU), *Game Design Patterns for Learning.*
**2012-42** Dominique Verpoorten (OU), *Reflection Amplifiers in self-regulated Learning.*
**2012-43** Withdrawn, .
**2012-44** Anna Tordai (VU), *On Combining Alignment Techniques.*
**2012-45** Benedikt Kratz (UvT), *A Model and Language for Business-aware Transactions.*
**2012-46** Simon Carter (UVA), *Exploration and Exploitation of Multilingual Data for Statistical Machine Translation.*
**2012-47** Manos Tsagkias (UVA), *Mining Social Media: Tracking Content and Predicting Behavior.*
**2012-48** Jorn Bakker (TUE), *Handling Abrupt Changes in Evolving Time-series Data.*
**2012-49** Michael Kaisers (UM), *Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions.*
**2012-50** Steven van Kervel (TUD), *Ontologogy driven Enterprise Information Systems Engineering.*
**2012-51** Jeroen de Jong (TUD), *Heuristics in Dynamic Sceduling; a practical framework with a case study in elevator dispatching.*

## 2013

**2013-01** Viorel Milea (EUR), *News Analytics for Financial Decision Support.*
**2013-02** Erietta Liarou (CWI), *MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing.*
**2013-03** Szymon Klarman (VU), *Reasoning with Contexts in Description Logics.*
**2013-04** Chetan Yadati(TUD), *Coordinating autonomous planning and scheduling.*
**2013-05** Dulce Pumareja (UT), *Groupware Requirements Evolutions Patterns.*
**2013-06** Romulo Goncalves(CWI), *The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience.*
**2013-07** Giel van Lankveld (UvT), *Quantifying Individual Player Differences.*
**2013-08** Robbert-Jan Merk(VU), *Making enemies: cognitive modeling for opponent agents in fighter pilot simulators.*
**2013-09** Fabio Gori (RUN), *Metagenomic Data Analysis: Computational Methods*

*and Applications.*
**2013-10** Jeewanie Jayasinghe Arachchige(UvT), *A Unified Modeling Framework for Service Design..*
**2013-11** Evangelos Pournaras(TUD), *Multi-level Reconfigurable Self-organization in Overlay Services.*
**2013-12** Marian Razavian(VU), *Knowledge-driven Migration to Services.*
**2013-13** Mohammad Safiri(UT), *Service Tailoring: User-centric creation of integrated IT-based homecare services to support independent living of elderly.*
**2013-14** Jafar Tanha (UVA), *Ensemble Approaches to Semi-Supervised Learning Learning.*
**2013-15** Daniel Hennes (UM), *Multiagent Learning - Dynamic Games and Applications.*
**2013-16** Eric Kok (UU), *Exploring the practical benefits of argumentation in multi-agent deliberation.*
**2013-17** Koen Kok (VU), *The PowerMatcher: Smart Coordination for the Smart Electricity Grid.*
**2013-18** Jeroen Janssens (UvT), *Outlier Selection and One-Class Classification.*
**2013-19** Renze Steenhuizen (TUD), *Coordinated Multi-Agent Planning and Scheduling.*
**2013-20** Katja Hofmann (UvA), *Fast and Reliable Online Learning to Rank for Information Retrieval.*
**2013-21** Sander Wubben (UvT), *Text-to-text generation by monolingual machine translation.*
**2013-22** Tom Claassen (RUN), *Causal Discovery and Logic.*
**2013-23** Patricio de Alencar Silva(UvT), *Value Activity Monitoring.*
**2013-24** Haitham Bou Ammar (UM), *Automated Transfer in Reinforcement Learning.*
**2013-25** Agnieszka Anna Latoszek-Berendsen (UM), *Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System.*
**2013-26** Alireza Zarghami (UT), *Architectural Support for Dynamic Homecare Service Provisioning.*
**2013-27** Mohammad Huq (UT), *Inference-based Framework Managing Data Provenance.*
**2013-28** Frans van der Sluis (UT), *When Complexity becomes Interesting: An Inquiry into the Information eXperience.*
**2013-29** Iwan de Kok (UT), *Listening Heads.*
**2013-30** Joyce Nakatumba (TUE), *Resource-Aware Business Process Management: Analysis and Support.*
**2013-31** Dinh Khoa Nguyen (UvT), *Blueprint Model and Language for Engineering Cloud Applications.*
**2013-32** Kamakshi Rajagopal (OUN), *Networking For Learning; The role of Networking in a Lifelong Learner's Professional Development.*
**2013-33** Qi Gao (TUD), *User Modeling and Personalization in the Microblogging*

*Sphere.*
**2013-34** Kien Tjin-Kam-Jet (UT), *Distributed Deep Web Search.*
**2013-35** Abdallah El Ali (UvA), *Minimal Mobile Human Computer Interaction.*
**2013-36** Than Lam Hoang (TUe), *Pattern Mining in Data Streams.*
**2013-37** Dirk Brner (OUN), *Ambient Learning Displays.*
**2013-38** Eelco den Heijer (VU), *Autonomous Evolutionary Art.*
**2013-39** Joop de Jong (TUD), *A Method for Enterprise Ontology based Design of Enterprise Information Systems.*
**2013-40** Pim Nijssen (UM), *Monte-Carlo Tree Search for Multi-Player Games.*
**2013-41** Jochem Liem (UVA), *Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning.*
**2013-42** Lon Planken (TUD), *Algorithms for Simple Temporal Reasoning.*
**2013-43** Marc Bron (UVA), *Exploration and Contextualization through Interaction and Concepts.*

## 2014

**2014-01** Nicola Barile (UU), *Studies in Learning Monotone Models from Data.*
**2014-02** Fiona Tuliyano (RUN), *Combining System Dynamics with a Domain Modeling Method.*
**2014-03** Sergio Raul Duarte Torres (UT), *Information Retrieval for Children: Search Behavior and Solutions.*
**2014-04** Hanna Jochmann-Mannak (UT), *Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation.*
**2014-05** Jurriaan van Reijsen (UU), *Knowledge Perspectives on Advancing Dynamic Capability.*
**2014-06** Damian Tamburri (VU), *Supporting Networked Software Development.*
**2014-07** Arya Adriansyah (TUE), *Aligning Observed and Modeled Behavior.*
**2014-08** Samur Araujo (TUD), *Data Integration over Distributed and Heterogeneous Data Endpoints.*
**2014-09** Philip Jackson (UvT), *Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language.*
**2014-10** Ivan Salvador Razo Zapata (VU), *Service Value Networks.*
**2014-11** Janneke van der Zwaan (TUD), *An Empathic Virtual Buddy for Social Support.*
**2014-12** Willem van Willigen (VU), *Look Ma, No Hands: Aspects of Autonomous Vehicle Control.*
**2014-13** Arlette van Wissen (VU), *Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains.*
**2014-14** Yangyang Shi (TUD), *Language Models With Meta-information.*
**2014-15** Natalya Mogles (VU), *Agent-Based Analysis and Support of Human Func-*

*tioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare.*
**2014-16** Krystyna Milian (VU), *Supporting trial recruitment and design by automatically interpreting eligibility criteria.*
**2014-17** Kathrin Dentler (VU), *Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability.*
**2014-18** Mattijs Ghijsen (VU), *Methods and Models for the Design and Study of Dynamic Agent Organizations.*
**2014-19** Vincius Ramos (TUE), *Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support.*
**2014-20** Mena Habib (UT), *Named Entity Extraction and Disambiguation for Informal Text: The Missing Link.*
**2014-21** Kassidy Clark (TUD), *Negotiation and Monitoring in Open Environments.*
**2014-22** Marieke Peeters (UU), *Personalized Educational Games - Developing agent-supported scenario-based training.*
**2014-23** Eleftherios Sidirourgos (UvA/CWI), *Space Efficient Indexes for the Big Data Era.*
**2014-24** Davide Ceolin (VU), *Trusting Semi-structured Web Data.*
**2014-25** Martijn Lappenschaar (RUN), *New network models for the analysis of disease interaction.*
**2014-26** Tim Baarslag (TUD), *What to Bid and When to Stop.*
**2014-27** Rui Jorge Almeida (EUR), *Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty.*
**2014-28** Anna Chmielowiec (VU), *Decentralized k-Clique Matching.*
**2014-29** Jaap Kabbedijk (UU), *Variability in Multi-Tenant Enterprise Software.*
**2014-30** Peter de Cock (UvT), *Anticipating Criminal Behaviour.*
**2014-31** Leo van Moergestel (UU), *Agent Technology in Agile Multiparallel Manufacturing and Product Support.*
**2014-32** Naser Ayat (UvA), *On Entity Resolution in Probabilistic Data.*
**2014-33** Tesfa Tegegne (RUN), *Service Discovery in eHealth.*
**2014-34** Christina Manteli(VU), *The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems..*
**2014-35** Joost van Ooijen (UU), *Cognitive Agents in Virtual Worlds: A Middleware Design Approach.*
**2014-36** Joos Buijs (TUE), *Flexible Evolutionary Algorithms for Mining Structured Process Models.*
**2014-37** Maral Dadvar (UT), *Experts and Machines United Against Cyberbullying.*
**2014-38** Danny Plass-Oude Bos (UT), *Making brain-computer interfaces better: improving usability through post-processing..*
**2014-39** Jasmina Maric (UvT), *Web Communities, Immigration, and Social Capital.*
**2014-40** Walter Omona (RUN), *A Framework for Knowledge Management Using ICT in Higher Education.*
**2014-41** Frederic Hogenboom (EUR), *Automated Detection of Financial Events in*

*News Text.*
**2014-42** Carsten Eijckhof (CWI/TUD), *Contextual Multidimensional Relevance Models.*
**2014-43** Kevin Vlaanderen (UU), *Supporting Process Improvement using Method Increments.*
**2014-44** Paulien Meesters (UvT), *Intelligent Blauw. Met als ondertitel: Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden..*
**2014-45** Birgit Schmitz (OUN), *Mobile Games for Learning: A Pattern-Based Approach.*
**2014-46** Ke Tao (TUD), *Social Web Data Analytics: Relevance, Redundancy, Diversity.*
**2014-47** Shangsong Liang (UVA), *Fusion and Diversification in Information Retrieval.*

## 2015

**2015-01** Niels Netten (UvA), *Machine Learning for Relevance of Information in Crisis Response.*
**2015-02** Faiza Bukhsh (UvT), *Smart auditing: Innovative Compliance Checking in Customs Controls.*
**2015-03** Twan van Laarhoven (RUN), *Machine learning for network data.*
**2015-04** Howard Spoelstra (OUN), *Collaborations in Open Learning Environments.*
**2015-05** Christoph Bsch(UT), *Cryptographically Enforced Search Pattern Hiding.*
**2015-06** Farideh Heidari (TUD), *Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes.*
**2015-07** Maria-Hendrike Peetz(UvA), *Time-Aware Online Reputation Analysis.*
**2015-08** Jie Jiang (TUD), *Organizational Compliance: An agent-based model for designing and evaluating organizational interactions.*
**2015-09** Randy Klaassen(UT), *HCI Perspectives on Behavior Change Support Systems.*
**2015-10** Henry Hermans (OUN), *OpenU: design of an integrated system to support lifelong learning.*
**2015-11** Yongming Luo (TUE), *Designing algorithms for big graph datasets: A study of computing bisimulation and joins.*
**2015-12** Julie M. Birkholz (VU), *Modi Operandi of Social Networks: The Effect of Context on Scientific Collaboration Networks.*